

HOW JURORS' BELIEFS COUNT

*Jack H.L. Whiteley**

INTRODUCTION	384
<i>Problem 1. Statistical Evidence</i>	385
<i>Problem 2. The Conjunction of Elements</i>	387
I. THE CREDENCE EXPLANATION	394
A. <i>The Credence Explanation</i>	394
B. <i>Its Problems and Proposed Solutions</i>	398
1. Statistical Evidence.....	398
a. <i>Missing Evidence</i>	398
b. <i>Social Incentives</i>	401
c. <i>Reference Class Problems</i>	403
d. <i>Public Acceptance</i>	406
2. The Conjunction of Elements	408
a. <i>Rarity</i>	408
b. <i>A Judicial Mistake</i>	411
c. <i>Public Acceptance</i>	412
II. OTHER ACCOUNTS OF THE STANDARDS OF PROOF	413
A. <i>Relative Plausibility Judgments</i>	414
B. <i>Weight</i>	419
C. <i>Probabilistic Knowledge</i>	423
III. THE OUTRIGHT BELIEF REQUIREMENT	426
A. <i>People Have Outright Beliefs</i>	428
B. <i>People's Willingness to Form Outright Beliefs</i> <i>Depends on the Circumstances in Which They</i> <i>Are Making Up Their Minds</i>	430
C. <i>Operationalizing the Outright Belief</i> <i>Requirement</i>	436

* J.D., Yale Law School, 2018. I am grateful to Lea Brilmayer, Catherine Chen, Richard Chen, Daniel Greco, Nicholas McLean, Jeessoo Nam, Kelley Schiffman, and two unknown readers for their thoughtful comments on previous drafts. Many thanks also to Jordan Thomas, Elena Peters, and the other student editors of the *Mississippi Law Journal*. All remaining errors are my own.

CONCLUSION.....	439
-----------------	-----

INTRODUCTION

There is a mystery in the standards of trial proof. In criminal trials, the jury is instructed:¹

[Y]ou must find the defendant not guilty, unless, on the evidence presented at this trial, you conclude that the People have proven the defendant guilty *beyond a reasonable doubt*.²

In civil trials, the jury is generally instructed:

[T]he plaintiff has the burden of proving the . . . elements [of the claim] by *a preponderance of the evidence*.³

For centuries,⁴ much has turned on these standards. But what do they mean?

On the conventional view, they tell the jurors the *degree* of certainty required to find the defendant liable or guilty. They establish the *credences*, or probability assessments, that must exist in the jurors' minds if civil or criminal liability is to ensue. The preponderance standard requires that jurors think that it is more than 0.5, or 50%, likely that the defendant did what the plaintiff says he has done.⁵ The reasonable doubt standard requires

¹ With the exception of bench trials, in which there is no jury and the judge is the factfinder. The arguments in this Article apply to all factfinders, whether judge or jury, but for convenience, I refer to the factfinders as jurors throughout.

² See, e.g., N.Y. COURTS, CRIMINAL JURY INSTRUCTIONS & MODEL COLLOQUIES, FINAL INSTRUCTIONS at 7 (2018) (emphasis added), http://www.nycourts.gov/judges/cji/5-SampleCharges/CJI2d.Final_Instructions.pdf.

³ See, e.g., NINTH CIRCUIT JURY INSTRUCTIONS COMM., MANUAL OF MODEL CIVIL JURY INSTRUCTIONS at 100 (2017) (emphasis added), https://www.ce9.uscourts.gov/jury-instructions/sites/default/files/WPD/Civil_Instructions_2021_3%20.pdf [<https://perma.cc/7Q6T-TNMG>].

⁴ See John Leubsdorf, *The Surprising History of the Preponderance Standard of Civil Proof*, 67 FLA. L. REV. 1569, 1569 (2015) (observing that judges started to instruct juries on the preponderance standard “in the mid-nineteenth century”); Barbara Shapiro, *Changing Language, Unchanging Standard: From ‘Satisfied Conscience’ to ‘Moral Certainty’ and ‘Beyond Reasonable Doubt,’* 17 CARDOZO J. INT’L & COMP. L. 261, 274 (2009) (noting that the “[b]eyond reasonable doubt language was clearly voiced in the Boston Massacre trials of 1770.”).

⁵ See, e.g., *Brown v. Bowen*, 847 F.2d 342, 345 (7th Cir. 1988) (“[T]he trier of fact rules for the plaintiff if it thinks the chance greater than 0.5 that the plaintiff is in the

something higher—perhaps 0.8, or 0.9, or 0.99, some other credence that approaches, but never reaches, 1.⁶

The credence account's appeal derives from how simply it models the distinction between the criminal and civil standards of proof. It is easier to find a civil defendant liable than to convict a criminal defendant, and the credence account seems to explain this difference in unimpeachably quantified terms.

But two problems have long troubled it. The first involves statistical evidence, and the second involves cases in which more than one element of a claim is disputed.

Problem 1. Statistical Evidence

Imagine a person is driving on a two-lane road on a moonless night. Suddenly, she sees the headlights of a bus careening toward her and she is forced to swerve into a ditch to avoid it, damaging her car in the process. Everything happened so quickly that she was not able to see very much. She knows only that it was a bus. Suppose she goes to court and presents this testimony, along with undisputed evidence that the defendant, the Blue Bus Company, operates 80% of all the buses on this road.⁷

Conceding both facts, the Blue Bus Company moves for summary judgment, arguing that there is no dispute as to any material fact and the company is entitled to judgment as a matter

right.”); *Althen v. Sec’y of Dep’t of Health & Human Servs.*, 58 Fed. Cl. 270, 283 (2003) (explaining that judges often say “the plaintiff must establish the facts necessary to her [or his] case by a probability greater than 0.5 or greater than 50%”) (quoting Dan B. Dobbs, *THE LAW OF TORTS* 360 (2000)); see also NINTH CIRCUIT JURY INSTRUCTIONS COMM., *supra* note 3, at 8 (“When a party has the burden of proving any claim [or affirmative defense] by a preponderance of the evidence, it means you must be persuaded by the evidence that the claim [or affirmative defense] is more probably true than not true.”).

⁶ *Brown*, 847 F.2d at 345-46 (“The reasonable doubt standard is much higher, perhaps 0.9 or better.”); *United States v. Fatico*, 458 F. Supp. 388, 410 (E.D.N.Y. 1978), *aff’d*, 603 F.2d 1053 (2d Cir. 1979) (reporting results of a survey of federal judges in the Eastern District of New York, who generally found the reasonable doubt standard was satisfied by probabilities between 0.8 and 0.9).

⁷ This version of the Blue Bus problem is adapted from Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 *YALE L.J.* 1254, 1269 (2013). The problem was originally developed by Laurence H. Tribe in *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 *HARV. L. REV.* 1329, 1340-41 (1971); see also Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 *HARV. L. REV.* 1357, 1378-79 (1985).

of law.⁸ The case is governed by the preponderance standard. Should this claim get to a jury?

Most readers, as well as courts, think not.⁹ The judge must award summary judgment to the Blue Bus Company. But why should this be? The plaintiff has presented evidence that there is a 0.8 chance that the Blue Bus Company damaged her car, which is well above the 0.5 credence threshold that the preponderance standard is said to require. Indeed, if the defendant concedes the claim has been proven to a probability of 0.8, shouldn't summary judgment, if anything, go the other way—in favor of the plaintiff?

A similar set of facts may be imagined for the reasonable doubt standard. Suppose that 0.95 is the credence the criminal standard requires.

In an enclosed yard are twenty-five identically dressed prisoners and a prison guard. The sole witness is too far away to distinguish individual features. He sees the guard, recognizable by his uniform, trip and fall, apparently knocking himself out. The prisoners huddle and argue. One breaks away

⁸ See, e.g., 10A CHARLES A. WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 2711 (4th ed. 2008), Westlaw (database updated Oct. 2020) (describing the summary judgment standard for federal courts).

⁹ See, e.g., Hock Lai Ho, *The Legal Concept of Evidence*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY § 3.2.2 (Edward N. Zalta ed. 2015), <https://plato.stanford.edu/entries/evidence-legal/#ObjUsiMatProIntStaPro> [<https://perma.cc/3LZE-KWHE>] (“[A]ll lawyers would agree that the evidence [in Blue Bus] is insufficient.”); *Howard v. Wal-Mart Stores, Inc.*, 160 F.3d 358, 360 (7th Cir. 1998) (“The plaintiff . . . asks for judgment on the basis of [statistical evidence] alone If the defendant also puts in no evidence, should a jury be allowed to award judgment to the plaintiff? The law’s answer is no.”); *Spencer v. Baxter Int’l, Inc.*, 163 F. Supp. 2d 74, 80 n.7 (D. Mass. 2001) (finding statistical evidence alone is not enough to avoid summary judgment); *Baker v. Bridgestone/Firestone Co.*, 966 F. Supp. 874, 876 (W.D. Mo. 1996). The court that considered perhaps the closest real-life case to this hypothetical agreed. In *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754, 755 (Mass. 1945), the court held 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.” See also Tribe, *supra* note 7, at 1341 n.37. A similar case, *Kaminsky v. Hertz Corp.*, 288 N.W.2d 426, 427 (Mich. Ct. App. 1979), held that such market-share evidence could, in conjunction with unchallenged eyewitness testimony that the vehicle bore a firm’s logo, establish a presumption that that firm owned the vehicle in question. For more on these cases, see Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1508 n.68 (1999), and David Enoch & Talia Fisher, *Sense and “Sensitivity”: Epistemic and Instrumental Approaches to Statistical Evidence*, 67 STAN. L. REV. 557, 561-62 (2015).

from the others and goes to a shed in the corner of the yard to hide. The other twenty-four set upon the fallen guard and kill him. After the killing, the hidden prisoner emerges from the shed and mixes with the other prisoners. When the authorities later enter the yard, they find the dead guard and the twenty-five prisoners. Given these facts, twenty-four of the twenty-five are guilty of murder.¹⁰

Were a prosecutor to choose one of these prisoners at random and try him for murder, there would be a 0.96 likelihood that he is guilty. And yet our intuitions rebel against conviction, just as any court would surely grant a motion for acquittal.

Why is this, if all that is required for guilt is a 0.95 likelihood? On its face, the credence account provides no answer.¹¹

Problem 2. The Conjunction of Elements

A separate problem arises with claims that have more than one disputed element. Imagine that a court is adjudicating a claim with three elements. The plaintiff must prove that: (1) the defendant was driving negligently; (2) the defendant's negligence caused a crash with the plaintiff's car; and (3) the plaintiff suffered an injury as a result.¹² By the end of the trial, the jury concludes the plaintiff has proven each of these elements to a probability of 0.6. There is a 60% chance that the defendant was driving negligently; a 60% chance that the negligence caused the crash; and a 60% chance that the plaintiff suffered a resultant injury. Has the plaintiff won?

The plaintiff has surpassed the 0.5 threshold on each element. But he may not have surpassed the 0.5 threshold in showing *all* the

¹⁰ Charles R. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1192-93 (1979); see also SARAH MOSS, *PROBABILISTIC KNOWLEDGE* 204-05 (2018).

¹¹ For a more detailed discussion of statistical evidence cases, see Mike Redmayne, *Exploring the Proof Paradoxes*, 14 LEGAL THEORY 281 (2008).

¹² This version of the paradox is adapted from Cheng, *supra* note 7, at 1256. The paradox may have been originally developed in L. JONATHAN COHEN, *THE PROBABLE AND THE PROVABLE* 66 (1977). For more on its history, see Leubsdorf, *supra* note 4, at 1579 n.45 (tracing its origins to JEROME MICHAEL & MORTIMER J. ADLER, *THE NATURE OF JUDICIAL PROOF: AN INQUIRY INTO THE LOGICAL, LEGAL, AND EMPIRICAL ASPECTS OF THE LAW OF EVIDENCE* 141-42 (1931)).

elements are satisfied. To find the likelihood of three formally independent elements being present at the same time requires multiplying their respective probabilities.¹³ $0.6 \times 0.6 \times 0.6$ equals 0.216, so the plaintiff should lose, if the elements are independent.¹⁴ In this case, the elements are not formally independent: the defendant's negligence cannot have caused the crash unless the defendant was driving negligently.¹⁵ Most legal claims will have a similar structure.¹⁶ But even so, the probability that *all* the elements are satisfied will be lower than the probability that any individual element is satisfied. A defendant might have been driving negligently, yet still not have caused the claimed injury. Thus, the more elements a claim has, the higher the credence needed in each element for the plaintiff to win. The same is true for crimes. The more elements to an offense, the higher the factfinder's credence in each element of a crime must be to form the required credence—whatever that credence is—that the defendant is guilty of the crime.

The problem is that this result does not match the law. A criminal defendant is only entitled “to a jury determination that [he] is guilty of *every* element of the crime with which he is charged, beyond a reasonable doubt,”¹⁷ not all of them, or their conjunction. Civil jury instructions commonly indicate that plaintiffs must “prove *every* essential part of [their] claims by a preponderance of the evidence,” and not, by implication, their conjunction.¹⁸ If courts required proof of the conjunction of the elements, then special

¹³ See, e.g., Cheng, *supra* note 7, at 1256-57.

¹⁴ See David S. Schwartz & Elliott Sober, *The Conjunction Problem and the Logic of Jury Findings*, 59 WM. & MARY L. REV. 619, 629 (2017).

¹⁵ See Carl G. Wagner, *The Probable and the Provable by L. Jonathan Cohen*, 1979 DUKE L. J. 1071, 1076 (1979) (book review); Schwartz & Sober, *supra* note 14, at 655-68.

¹⁶ See *infra* Part I.

¹⁷ *Apprendi v. New Jersey*, 530 U.S. 466, 477 (2000) (quoting *United States v. Gaudin*, 515 U.S. 506, 510 (1995)) (emphasis added).

¹⁸ See, e.g., JUDICIAL COUNCIL OF THE UNITED STATES COURT OF APPEALS FOR THE ELEVENTH CIRCUIT, ELEVENTH CIRCUIT PATTERN JURY INSTRUCTIONS (CIVIL CASES) § 3.7.1 (2020) (emphasis added), <http://www.ca11.uscourts.gov/sites/default/files/courtdocs/clk/FormCivilPatternJuryInstructionsCurrentComplete.pdf?revDate=20190124> [https://perma.cc/22FB-8R37]; see also Schwartz & Sober, *supra* note 14, at 673-74 (finding that fewer than one-quarter of U.S. jurisdictions have such civil jury instructions, one-third have such criminal jury instructions, and that those that do not have such instructions nonetheless permit special verdict forms).

verdicts, a relatively common practice which requires jurors to make binary yes/no assessments of each element to be proved, would be out of the question.¹⁹ If credences below 1 really are what the standards of proof require, then proof of each element is not the same as proof of the whole claim, and every real-world case like this negligence hypothetical risks causing an unjust verdict.

The statistical evidence and conjunction problems are structured in the same way, presenting a case that the credence account is unable to resolve in a way that is consistent with what the law requires. In their complexity, they have influenced fruitful scholarship about the assumptions, goals, and policies underlying the law,²⁰ general questions about the value of mathematics in trial evidence regimes,²¹ and empirical examinations of how jurors resolve cases in practice.²²

But they raise above all a narrow, everyday question of legal interpretation. What is required for a jury to find that a claim has been proven? The problems reveal enduring mysteries in the words “beyond a reasonable doubt” and “by a preponderance of the evidence.” They whisper that we do not know what these commands mean, or what jurors are supposed to do with them. This is troubling, because they are the commands that determine whether people are found to have broken the law, and so made to pay

¹⁹ See, e.g., Kevin M. Clermont, *Death of Paradox: The Killer Logic Beneath the Standards of Proof*, 88 NOTRE DAME L. REV. 1061, 1110 (2013). Special verdict forms, which trial counsel may request, and courts may require, are written questions to the jury on each contested issue of fact that are “susceptible of a categorical or other brief answer.” See FED. R. CIV. P. 49(a)(1)(A); see also 6 AM. JUR. TRIALS 1043, Westlaw (database updated Dec. 2020) (“The special verdict can often be a useful technique, and should at least be given consideration by trial counsel in every important case.”).

²⁰ See, e.g., MICHAEL O. FINKELSTEIN, *QUANTITATIVE METHODS IN LAW* (1978); Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116 (1978) (reviewing MICHAEL O. FINKELSTEIN, *QUANTITATIVE METHODS IN LAW* (1978)); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34 (1979).

²¹ See, e.g., Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491 (2001); Dale A. Nance, *A Comment on the Supposed Paradoxes of a Mathematical Interpretation of the Logic of Trials*, 66 B.U. L. REV. 947, 947 (1986) (“[T]he debate reflects a more general controversy over the virtues of quantifying legal decisionmaking.”).

²² See, e.g., Nancy Pennington & Reid Hastie, *A Cognitive Theory of Juror Decision Making: The Story Model*, 13 CARDOZO L. REV. 519 (1991).

recompense or sent to prison. It would be a grave injustice if the standards of proof were incoherent.

This Article develops a conceptual solution. Each standard requires the jurors to form not a credence but an *outright belief* that the defendant has broken the law.²³ An outright belief is a belief to which a person does not consciously assign any likelihood but 1, or 100%. It feels like knowledge to the person who holds it.²⁴ It is an attitude of “settling on the truth” of a proposition.²⁵

This definition of outright belief will require elaboration, as well as a defense, which the Article will make in Part III, but it is intended to demonstrate the idea. If an outright belief, rather than any credence below 1, is what the law demands of jurors if they are to find a defendant liable or guilty, then the statistical evidence and conjunction puzzles are solved.

What is wrong with statistical evidence? As the philosopher Lara Buchak has argued,²⁶ it cannot cause a reasonable juror to believe outright that the defendant broke the law. At best, the evidence in *Blue Bus* can induce a credence in a reasonable juror of 0.8 that the bus company is responsible—not an outright belief to

²³ Some scholars conceive of the question as being what “legal proof” or “judicial proof” requires. See, e.g., MOSS, *supra* note 10, at 202; Nesson, *supra* note 7, at 1357. This Article’s ambition is narrower. It gives an account of what legal proof requires under the *reasonable doubt* and *preponderance* standards, and, by extension, intermediate civil standards like *clear and convincing evidence*. See *infra* Part III. Proof takes many forms in the American legal system, and there are some cases that are ungoverned by these standards. See, for example, claims in federal immigration court under the Convention Against Torture (“CAT”), which forbids signatory states from deporting people to countries in which they will be tortured upon return. See U.N. G.A. Res. 39/46, annex, Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment (Dec. 10, 1984). An alien charged with deportation may make a claim under the CAT, arguing that that he will face torture if returned to his homeland. Under federal regulations, to obtain relief, the claimant must show simply that he is “more likely than not” to face torture if he is returned. 8 C.F.R. § 208.16 (2021). Legal proof under the CAT, then, facially requires only a credence greater than 0.5 that the claimant will face torture if returned. CAT cases also do not seem to be susceptible to statistical evidence problems: if a CAT claimant credibly showed that his home government subjects 80% of all people in the country to torture, then this certainly would permit relief under the statute. This is one example, and there are likely others.

²⁴ Daniel Greco, *How I Learned to Stop Worrying and Love Probability 1*, 29 PHILOSOPHICAL PERSPECTIVES 179, 180 (2015).

²⁵ Jacob Ross & Mark Schroeder, *Belief, Credence, and Pragmatic Encroachment*, 88 PHIL. & PHENOMENOLOGICAL RESEARCH 259, 286 (2014).

²⁶ Lara Buchak, *Belief, Credence, and Norms*, 169 PHILOSOPHICAL STUDIES 285, 308-09 (2013). See also *infra* Part III.

that effect. Prison Yard is the same. The evidence induces a credence of at best 0.96 that the randomly selected prisoner took part in the murder. In neither case can a reasonable person form an outright belief—a credence of 1—that the defendant broke the law based on the statistical evidence alone. And that is statistical evidence's peculiar weakness.

Other kinds of evidence, like a scratch of blue paint or an eyewitness who testifies she saw the bus's logo clearly, *can* cause a reasonable person to believe outright that a particular bus injured someone. This is true even though such kinds of evidence are not perfectly reliable, and even if the believer knows that.²⁷ Even if she knows, for example, that an eyewitness is generally reliable only 80% of the time. Or so this Article will argue.

By contrast, a reasonable juror who holds the Blue Bus company liable based on the statistical evidence alone must consciously be gambling. She must know there is a 0.2 chance that the defendant is innocent. Without being able to rule out this chance in her own mind, all she can do is hope that the gamble is not unlucky.²⁸ The same is true of Prison Yard, a gamble with better odds, but a gamble still.²⁹

There is also no conjunction paradox if an outright belief—a credence of 1—is required for guilt or liability. If a claim has more than one element, then each element must always be proven to a credence of 1 for the whole claim to be proven to a credence of 1. If a claim has three elements, anything less than $1 \times 1 \times 1$ cannot equal 1. Indeed, 1—outright belief—is the only credence that never

²⁷ The gap between the reliability and the persuasive power of eyewitness testimony in particular has caused concern among lawyers and judges. *See, e.g.*, Jed S. Rakoff, *Our Lying Eyes*, N.Y. REV. BOOKS (Apr. 18, 2019), <https://www.nybooks.com/articles/2019/04/18/our-lying-eyes/> [<https://perma.cc/9MRY-VTXM>].

²⁸ For other consideration of possible connections between statistical evidence and luck, see Judith Jarvis Thomson, *Liability and Individualized Evidence*, 49 *LAW & CONTEMP. PROBS.* 199, 201-11 (1986).

²⁹ The Article uses “an outright belief that the defendant broke the law” as a shorthand because of the usual procedural posture of the burdens of proof. What the Article means more precisely is, “an outright belief that what is claimed to have happened in fact happened.” For example, when the procedural posture is such that a defendant must prove an affirmative defense by a preponderance of the evidence, then the juror must form an outright belief that what the defendant claims has happened did in fact happen to sustain a finding that supports that affirmative defense.

results in a conjunction paradox. It is the only credence for which proof of every element is the same as proof of all the elements.

Yet, an outright belief requirement has rarely been contemplated,³⁰ and two barriers to it come to mind. First, it seems to collapse the distinction between the different standards of proof. If the law requires jurors to form a credence of 1 under each standard, then there appears to be no difference between the preponderance and the reasonable doubt standards. Yet the two standards clearly are different. Everyone agrees that it is, and should be, harder to convict a criminal defendant than to hold liable a civil defendant.

Second, an outright belief requirement might be too exacting for *any* standard of trial proof. A credence of 1 sounds like a rigid and immutable certainty, and if the law required that kind of certainty, then it might demand too many acquittals. In our uncertain world in which it is hard to rule out all kinds of possibilities, it may be imprudent to have that kind of certainty in much of anything, and if this were so, finding anyone guilty might be imprudent. This objection is even more powerful when applied to the preponderance standard, which is weaker than the criminal standard.

³⁰ As Part III discusses, some legal scholars have argued for it in the criminal context. *See, e.g.,* Andrea Roth, *Safety in Numbers? Deciding When DNA Alone is Enough to Convict*, 85 N.Y.U. L. REV. 1130, 1184 (2010). In considering statistical evidence and its relation to belief, credence, and blame, Lara Buchak has notably argued that it applies in the civil context also. *See* Buchak, *supra* note 26, at 290-91. The same suggestion—that civil cases require outright belief—may have been made at the beginning of the twentieth century. *See* William Trickett, *Preponderance of Evidence, and Reasonable Doubt*, 10 FORUM 75, 78 (1906) (“Is not the principle abhorrent that B may be coerced into paying a sum of money to A, when the jury does not believe, even in a faint degree, that he promised to pay it, simply because it believes that, of the plaintiff’s and defendant’s respective pieces of evidence, that of the former is heavier than that of the latter?”).

Some courts have arguably suggested an outright belief is required in the civil context, although this is a matter of interpretation. *See* *Sargent v. Massachusetts Accident Co.*, 29 N.E.2d 825, 827 (Mass. 1940) (“After the evidence has been weighed, that proposition is proved by a preponderance of the evidence if it is made to appear more likely or probable in the sense that actual belief in its truth, derived from the evidence, exists in the mind or minds of the tribunal notwithstanding any doubts that may still linger there.”). What is the *Sargent* court asserting? On the one hand, the court suggests the tribunal needs to actually believe in its truth; on the other, it mentions that doubts may still linger.

The objections are at first compelling, but fifty-year-old problems call for creative thinking. And the objections are answered easily enough, so long as the law has a certain view of how jurors' minds work. First, outright beliefs must be real. People must be able to believe things outright, rather than just form high credences about them that are close to but below 1. Second, people's willingness to form outright beliefs must depend not just on the evidence available to us, but on the circumstances in which we find ourselves when we are making up our minds. If the same person, in different *circumstances*, can be less willing to form the *same* outright belief based on the *same* evidence, then the distinction between the standards of proof can be preserved.

Recent developments in the philosophy of knowledge make these commitments plausible. Thanks to people's limited cognitive capacities, we need outright beliefs to function. And, in part as a result of this, our willingness to form such beliefs depends on the circumstances in which we find ourselves when we are making up our minds.

Although this is a theoretical argument, it has a practical consequence, resolving a longstanding split among judges over how to instruct juries on the preponderance standard. Some judges' instructions say that the preponderance standard requires the jury to find the plaintiff's claim is "more probably true than not true,"³¹ while others eschew this probabilistic language, instead telling jurors to consider the "weight of the evidence,"³² or describing the preponderance standard with metaphors like "an old-fashioned balanc[ing] scale."³³ It is hard to interpret the "more probably true than not true" language to mean anything other than "you must believe the plaintiff's claim is at least 0.51 likely to be true," and so, if this Article is right, this instruction incorrectly describes the law. The metaphor instructions are to be preferred.

But why should the law have an outright belief requirement? What justifies it? An initial thought is that a person who believes something outright may be wrong, but she is not consciously

³¹ See, e.g., NINTH CIRCUIT JURY INSTRUCTIONS COMM., *supra* note 3, at 314.

³² See Leubsdorf, *supra* note 4, at 1572 (quoting the Virginia model jury instructions).

³³ See *id.* (quoting the Pennsylvania model civil jury instructions); see also *infra* Part III.

gambling on its truth.³⁴ If the law is to empower people to act as jurors, and so decide whether others have done things that will subject them to penalties and punishments, then an outright belief requirement is meant to demand that kind of certainty. It forbids juries from gambling with other people's fates.

Part I of the Article gives an overview of the credence account and the puzzles of the statistical evidence and conjunction cases, considers several of the most promising proposed solutions to them, and argues that they do not succeed.³⁵ Part II considers three of the most promising alternative accounts of the standards of proof, which aim to avoid the credence account's problems, and finds that they, also, are unsatisfactory. Part III contains the Article's affirmative argument: the standards of proof require outright belief.

I. THE CREDENCE EXPLANATION

A. *The Credence Explanation*

A standard of proof, Justice Harlan once held, "represents an attempt to instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication."³⁶ Although the standards "are quantitatively imprecise," they communicate to the jurors the degree of confidence they must have that the defendant broke the law.³⁷

This view has gained a great deal of recognition. It has been said that all first-year law students learn that the civil preponderance standard requires the plaintiff to establish her claim to a probability of 0.51 or higher.³⁸ A significant number of pattern jury instructions on the preponderance standard state that

³⁴ See Buchak, *supra* note 26, at 292-94.

³⁵ This is not an exhaustive account of the solutions that have been offered. The rich and cross-disciplinary inquiry that the proof paradoxes have caused cannot be adequately summarized here. My hope is simply to identify the solutions that I regard as the most promising.

³⁶ *In re Winship*, 397 U.S. 358, 370 (1970) (Harlan, J., concurring).

³⁷ *Id.*

³⁸ Cheng, *supra* note 7, at 1256 ("As every first-year law student knows, the civil preponderance-of-the-evidence standard requires that a plaintiff establish the probability of her claim to greater than 0.5.").

it is satisfied if the claim is more probably true than not true.³⁹ Even some judicial opinions have said so.⁴⁰ The idea is intuitive, and, at first glance, might seem irrefutable. To quote the U.S. Court of Appeals for the Seventh Circuit, “All evidence is probabilistic—statistical evidence merely explicitly so.”⁴¹

Things become more complicated when we consider the criminal standard, as the precise credence required for “beyond a reasonable doubt” is disputed. In an anonymous survey, almost a third of judges set it at 0.9 or 0.95, and a significant proportion set it as 0.8.⁴² These findings found their way into in an opinion from the federal district court for the Eastern District of New York, which considered the quantification question seriously.⁴³ Judge Jack Weinstein evidently went so far as to perform his own survey of judges in the Eastern District on how they might quantify the credence required for reasonable doubt, and found that they generally answered between 0.75 and 0.95.⁴⁴

The credence account’s most detailed conception is a variant of Bayesian decision theory first articulated by John Kaplan several decades ago.⁴⁵ On his view, the jurors must consider the evidence, weighing each item, and determine how it affects their view of the probabilities as a whole.⁴⁶ The process is, “for the decision theorist, quantified by Bayes’ Theorem.”⁴⁷ Does this item of evidence make it more or less likely that the defendant broke the law, and by how much?⁴⁸ Highly probative evidence will have a high likelihood ratio,

³⁹ See, e.g., NINTH CIRCUIT JURY INSTRUCTIONS COMM. , *supra* note 3, at 8 (stating that the preponderance standard requires that the jury find the plaintiff’s claim to be “more probably true than not true.”).

⁴⁰ See *supra* note 5.

⁴¹ Riordan v. Kempiners, 831 F.2d 690, 698 (7th Cir. 1987).

⁴² See Barbara D. Underwood, *The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases*, 86 YALE L.J. 1299, 1311 (1977) (citing Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom*, 5 L. & SOC’Y REV. 319, 325 (1971). Instructively, about a third set it as 1. *Id.*

⁴³ *United States v. Fatico*, 458 F. Supp. 388, 409-10 (E.D.N.Y. 1978). A single judge replied that the reasonable doubt standard could not be estimated numerically. *Id.* at 410.

⁴⁴ *Id.* at 410.

⁴⁵ See generally John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1968).

⁴⁶ *Id.* at 1083.

⁴⁷ *Id.*

⁴⁸ *Id.*

and minimally probative evidence will have a low one.⁴⁹ The jurors are tasked with considering the likelihood ratios of each piece of evidence in the case.⁵⁰ They “will evaluate each piece of evidence as it comes in, applying Bayes’ Theorem to examine whether and how much it aids the prosecution or the defense.”⁵¹

This is the Bayesian explanation of the credence account, and several of its tenets remain undisputed. For example, *probability* accounts of the standards of proof are widely understood to be *credence* accounts.⁵² Kaplan argued that probability accounts are forced to aim at subjective probability—assessments from the jurors’ perspective—rather than something like objective probability, a slippery concept.⁵³ The objective probability of “a coin coming up heads in a given throw would be the ratio of heads to total tosses of the coin as the number of identical tosses approached infinity.”⁵⁴ But what does one mean by similar, or identical, tosses? If the tosses were identical, they would all land heads. The problem is magnified in the context of trials, when “it is meaningless to speak of the probability of the defendant’s guilt in terms of the number of times he would be guilty in an infinite number of exactly similar cases” because no two cases are exactly similar.⁵⁵ The Supreme Court came to the same view, finding that the reasonable doubt standard impresses “upon the factfinder the need to reach a *subjective state* of near certitude of the guilt of the accused.”⁵⁶

Kaplan also showed that a subjective probability assessment could be expressed as a quantitatively informed guess or bet.

⁴⁹ *Id.* at 1083-84.

⁵⁰ *Id.* at 1083.

⁵¹ *Id.* at 1084.

⁵² See, e.g., Jordi Ferrer Beltrán, *Legal Proof and Fact Finders’ Beliefs*, 12 LEGAL THEORY 293, 296 n.5 (2006) (“[A] standard of proof is no more than a rule for decision-making that indicates the minimum amount of corroboration required for a hypothesis to be considered proven. For this reason, if the standard refers to some kind of belief held by a person, i.e., the trier, then the result as to whether the hypothesis is proven or not will depend on that belief being held.”).

⁵³ Kaplan, *supra* note 45, at 1066.

⁵⁴ *Id.*

⁵⁵ *Id.* This criticism is elaborated further in the “reference class problem” criticism of statistical evidence, discussed *infra*, Part I.

⁵⁶ *Jackson v. Virginia*, 443 U.S. 307, 315 (1979) (emphasis added).

Thus, we could take a well-shuffled deck of 100 cards numbered consecutively from 1 to 100 and ask the subject whether he considered the probability of a given event to be the same as the probability that the number 1 would appear in, say, the top 10, 20, 40, or 50 cards.⁵⁷

This is intuitive. A unique card in a deck of 100 has a probability of 0.01 of being drawn at random from the deck. If asked what one thought the chances were of the card being drawn, so long as one was sure it was a fair deck of 100 containing just this one unique card, one would presumably answer: 0.01, or 1%. And if asked the odds of drawing a card that was *not* this unique card, one would answer: 0.99, or 99%.

The Bayesian model thus packages credence-based intuitions into a theory. Since the discovery of the statistical evidence and conjunction paradoxes, it has arguably been more frequently cited as the dominant view than defended,⁵⁸ but versions of the credence account continue to appear in jury instructions, judges' opinions, and casebooks. It is at first hard to imagine what another explanation of the standards of proof might look like, and perhaps no other explanation thus far offered has been as straightforward to understand.⁵⁹

⁵⁷ Kaplan, *supra* note 45, at 1067. Moreover, "by using differing numbers of cards we could define any other probability to any required degree of exactitude." *Id.* In this way, Kaplan adds mathematical rigor to a far older idea. Jeremy Bentham observed that "the practice of wagering affords at the same time a proof of the existence, and a mode of expression or measurement" for such "quantities or degrees" of legal proof. JEREMY BENTHAM, *Rationale of Judicial Evidence, Specially Applied to English Practice*, in 6 THE WORKS OF JEREMY BENTHAM 223 (John Bowring ed. 1843). *See also* MOSS, *supra* note 10, at 203.

⁵⁸ See, for example, the arguments cited *infra*, Part II. *See also* SUSAN HAACK, EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 61-62 (2014); Lea Brilmayer, *Second-Order Evidence and Bayesian Logic*, 66 B.U. L. REV. 673, 674 (1986) ("Bayesian probability theory is founded upon a form of logic that has been recognized as inadequate since the turn of the century."). *But see* Schwartz & Sober, *supra* note 14, at 626; Brian Hedden & Mark Colyvan, *Legal Probabilism: A Qualified Defence*, 27 J. POL. PHIL. 1 (2019).

⁵⁹ *See infra* Part II.

B. Its Problems and Proposed Solutions

1. Statistical Evidence

Statistical evidence, however, presents a problem. Lawyers' intuitions, as well as some judicial opinions, find that evidence of the kind presented in *Blue Bus* and *Prison Yard* cannot satisfy either standard of proof. Such evidence may be relevant, and therefore admissible when combined with other types of evidence,⁶⁰ but most people do not think it is enough, on its own, to sustain liability. If this commonly held view is correct, then the standards of proof require something more than probability assessments.

The statistical evidence cases suggest that some sort of "individualized evidence" is required for legal proof.⁶¹ But what is individualized evidence, and what makes it required for legal proof? Why do the standards of proof, which on their face contain no language in support of individualized evidence—or against statistics—require individualized evidence when confronted with cases in which only statistical evidence can form the basis of the verdict?

Many solutions to the problem have been proposed, seeking to make an individualized evidence requirement consistent with the credence account. I consider four such types of solution here.

a. Missing Evidence

In a foundational article, Laurence Tribe argued that what is important about the hypotheticals is the evidence that is missing. In few statistical evidence cases, he claimed, "can the mathematical evidence, *taken alone and in the setting of a completed lawsuit*, establish the proposition to which it is directed with sufficient probative force to prevail."⁶² In the *Blue Bus* case, the statistical evidence may not be enough to generate a credence of 0.8, or even 0.5, that it was a bus that belonged to the *Blue Bus Company* that

⁶⁰ See, e.g., *Kaminsky v. Hertz Corp.*, 288 N.W.2d 426, 427 (Mich. Ct. App. 1979); *Roberson v. State*, 16 S.W.3d 156, 166 (Tx. Ct. App. 2000) (discussing DNA evidence). For further discussion of courts' treatment of the relevance of statistics, see Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 *JURIMETRICS* 373, 377 (2002).

⁶¹ Thomson, *supra* note 28, at 206.

⁶² Tribe, *supra* note 7, at 1349.

nearly crashed into the plaintiff's car. This is because the lack of other evidence in the case might give rise to a negative inference about the plaintiff's claim.

Are we to believe, asks this argument, that if the Blue Bus Company caused the accident, the plaintiff would have uncovered nothing else to prove it? No other witnesses who saw the bus's logo or license plate? No Blue Bus Company timetables showing that one of its buses would have been on this part of the road at the time of the accident? Such absences are relevant to a reader's assessment of the hypothetical. It might be shown that in many instances, statistical evidence, at least of the kind presented in the Blue Bus hypothetical, fails to generate credences above 0.5.

But the possibility of missing evidence does not solve the problem. If the absence of other evidence in the Blue Bus hypothetical causes people to form credences below 0.5 that the Blue Bus Company broke the law, this explains why a jury might not find the company liable. It does not explain why a judge should forbid the case from reaching a jury at all.⁶³ The jurors could consider the evidence in Blue Bus and reach a credence lower than 0.5 that the Blue Bus Company broke the law, and so rule against the plaintiff.⁶⁴ The possibly instructive absences of evidence in Blue Bus do not seem to *require* courts to forbid jurors from considering the evidence, and thus uniformly to find for the defendant. And yet readers' intuitions and judges' orders favor disposing of the Blue Bus case at summary judgment. As Charles Nesson pointed out, Tribe's "argument leads to the conclusion that the case should reach the jury, and the jury's verdict should be upheld, no matter which way it comes out."⁶⁵

Lurking in the background, for Tribe, there seems to be a worry that jurors are incapable of properly evaluating statistical evidence.⁶⁶ Of central concern to his article was *People v. Collins*, a

⁶³ See Nesson, *supra* note 7, at 1380 ("Tribe's argument explains why a court should refuse to grant a directed verdict to the plaintiff, but his analysis does not explain why the judge should throw the plaintiff out of court.").

⁶⁴ Studies of mock jurors have concluded that jurors are more suspicious of statistical evidence than eyewitness testimony. See Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSONALITY & SOC. PSYCHOL. 739, 740, 744, 746, 748 (1992).

⁶⁵ Nesson, *supra* note 7, at 1381.

⁶⁶ Tribe, *supra* note 7, at 1355, 1336 n.23, 1662-63.

California case in which a prosecutor successfully convicted defendants of robbery by combining invented data with the expert testimony of a college math professor.⁶⁷ The eyewitnesses in the case had seen very little, and could only testify as to generalities about the robbers, including their race and hair color, and the color of their getaway car.⁶⁸ After calling the math professor to the stand, the prosecutor asked the professor to assume base probabilities for each of the characteristics the eyewitnesses testified about, and then asked the professor to state the probability of their combination.⁶⁹ The witness explained the conjunction rule that this Article has previously discussed, on which “the probability of the joint occurrence of a number of mutually independent events is equal to the product of the individual probabilities that each of the events will occur,” and then said the odds of a couple who matched the characteristics identified was one in twelve million, assuming the prosecutor’s base probabilities were true.⁷⁰ Turning to the jury, the prosecutor then asserted that “in reality ‘the chances of anyone else besides these defendants being there, . . . having every similarity, . . . is somewhat like one in a billion.’”⁷¹

The problems with this were manifold. For one thing, the prosecutor had plucked the base rates for the assailants’ characteristics out of thin air. And even if the invented rates were right, a 1-in-12-million random match probability emphatically does not mean the chances were 1 in 12 million that these defendants were innocent. The crime took place in a large city, in an age of car and airplane travel.⁷² If there were, say, 24 million people who could have visited the city that day, then the odds of the defendants’ guilt presented by the random match probability would have been 1 in 2, or 0.5—well below what anyone thinks is enough to satisfy the beyond-a-reasonable-doubt standard. Still, the defendants were convicted.⁷³ Mathematics, the California Supreme

⁶⁷ *Id.* See also *People v. Collins*, 438 P.2d 33, 36-37 (Cal. 1968).

⁶⁸ *Id.* at 34.

⁶⁹ See *id.* at 36-37.

⁷⁰ *Id.* at 36.

⁷¹ *Id.* at 37.

⁷² *Id.* at 34.

⁷³ *Id.* at 33.

Court later wrote when it reversed the convictions, was “a veritable sorcerer in our computerized society.”⁷⁴

Still, the evidence in the Blue Bus and Prison Yard hypotheticals does not suffer from these problems. It is not invented, confusing, or in any obvious sense incorrect. Why should jurors be incapable of assessing it?

b. Social Incentives

Perhaps the statistical evidence problem can be solved by appealing to public policy or social incentives. Richard Posner famously made such an argument. He agreed with Tribe that the problem with the Blue Bus hypothetical was “the tacit assumption that the statistic concerning the ownership of the buses is the only evidence that the plaintiff can obtain,”⁷⁵ and concluded there were two reasons why a jury should not be permitted to decide a case based upon it.

The first is simple judicial economy: “A court should not expend any of its scarce resources of time and effort on a case until the plaintiff has conducted a sufficient search to indicate that an expenditure of public resources is reasonably likely to yield a significant social benefit.”⁷⁶

The second is societal efficiency. Posner argued that a rule that held the Blue Bus Company liable based on statistical evidence alone would mean, over time, that the company would be held liable for all unexplained accidents.⁷⁷ No plaintiffs would have to conduct investigations of their own; they could simply present the market share evidence to a court and win, causing the Blue Bus Company to pay for every bus accident, even those it did not cause. This argument follows Tribe’s, who suggested that the upshot of improperly weighing the glaring omission of other evidence in Blue Bus “would be a regime in which the company owning four-fifths of” the buses “would have to pay for *five-fifths* of all unexplained bus accidents—a result as inefficient as it is unfair.”⁷⁸ This would leave the Blue Bus company’s competitors, according to Posner, with “no

⁷⁴ *Id.*

⁷⁵ Posner, *supra* note 9, at 1509.

⁷⁶ *Id.*

⁷⁷ *See id.* at 1510.

⁷⁸ Tribe, *supra* note 7, at 1349–50.

incentive to be careful,” for they will not be held liable for accidents,⁷⁹ and in so burdening the Blue Bus Company, such a rule would subsidize the competitors, who, having no legal encouragement to be careful, would cause more accidents. Eventually, the Blue Bus Company, “having higher liability costs, will probably withdraw from the route; the rule on burden of proof will have created a monopoly!”⁸⁰

But, just like the missing evidence argument, the appeal to incentives assumes that, unless the judge prevents the case from getting to a jury, the Blue Bus Company will surely lose. Social incentives might be a good argument against granting a directed verdict to the *plaintiff* in the Blue Bus case. But the other alternative—sending the case to the jury—would leave a different pattern of verdicts. The jury might just as well find *against* the plaintiff, thinking, as Tribe and Posner do, that the paltriness of the evidence in the case casts doubt on her claim. Thus, even without a legal rule granting summary judgment to the defendant, the Blue Bus Company would not be held liable for all unexplained accidents, as some juries would be uncomfortable with finding the company liable based on the statistical evidence alone.

Moreover, even if courts did give a directed verdict to the plaintiff, this would not result in the Blue Bus Company being held liable in all cases. Instead, the functional burden of producing evidence would shift from the plaintiff to the defendant. To avoid adverse summary judgment, it would be up to the Blue Bus Company to uncover evidence that some other bus caused the accident. Perhaps the company could turn over its bus timetables, showing none of its buses would have been on that stretch of road at that time of night. Perhaps it could hire an investigator to inquire into the other bus companies. In cases in which the Blue Bus Company did not cause the accident, then it might be able to find exculpatory evidence and so avoid liability.

Thus, granting summary judgment is not necessary to prevent the Blue Bus Company from being held liable for all unexplained bus accidents. If courts categorically refuse to allow the Blue Bus case to reach the jury, this rule is efficient only if it is cheaper for

⁷⁹ Posner, *supra* note 9, at 1510.

⁸⁰ *Id.*

the plaintiff to locate and introduce additional evidence than it would be for the Blue Bus Company to do so. This is an empirical question. Without knowing the answer, all that is assured about the summary judgment rule is that it protects dominant players in the bus industry.

The incentives-based argument makes even less sense in the criminal context. The argument would proceed as follows. Imagine the point of view of a person in Prison Yard, choosing whether to take part in the murder. An incentives-based argument would say that if he knows that 24 other prisoners are already involved in the crime, then he will have no legal incentive not to involve himself in the murder also. For if he chooses to participate, then he will be convicted: 100% of the prisoners will have participated in the crime, so all of them will be convicted. And if he chooses not to, he will be convicted all the same: 96% of the prisoners will have participated in the crime, and he will have no exculpatory evidence to show he was the one who did not. Thus, only by refusing to permit statistical evidence verdicts can courts give him—a person deliberating about whether to commit a crime—a reason to obey the law.

But this is obviously not likely to be the case in real life. Even if jurors always convicted on statistical evidence alone, the practical incentives would only emerge when a rare set of planets were aligned. It would require people, who would otherwise desire to break the law, to know enough about trials and evidence to know juries always convict in statistical evidence cases. They must know that they will be tried in court, and that only statistical evidence is going to be available in their trials. They must know they will be unable to produce any exculpatory evidence to differentiate themselves from the statistics. And they must have the time and coolness of mind to think through all these things in making their choices. It is possible that the law forbids statistical-evidence-based verdicts just to deter people in these circumstances from lawbreaking. But this must be a very small group of people.

c. Reference Class Problems

Another criticism the credence account could make of statistical evidence is that it suffers from an imponderable known

as the reference class problem. Ronald Allen and Michael Pardo have critiqued the Blue Bus hypothetical on these grounds.⁸¹

Recall that the plaintiff's evidence is that the Blue Bus Company operates 80% of the buses on the *road* where the plaintiff alleges the accident occurred. It is easy for the mind to translate this to mean the "correct" base rate of Blue Bus Company buses is 80%. But suppose it turned out that on a smaller section of road—say, the length of a city block—where the plaintiff alleges the accident occurred, the Blue Bus Company operates just 40% of the buses. Buses belonging to other companies perhaps use this north-south block as a convenient place to turn along their east-west routes, though they do not remain on the street for long, and many of the Blue Bus routes along this road end before—or start after—this city block. This new statistical evidence does not contradict the plaintiff's evidence that the Blue Bus Company operates 80% of the buses on the road, because the road and the city block represent different *reference classes*. The different result between the two classes is explained by the different number of buses in each class. The *road* class contains a larger number of buses, most of which are owned by the Blue Bus Company, than the *city block* class, most of which are not.

"Each of the reference classes leads to a different inference about" whether the Blue Bus Company is likely liable, "and nothing determines the correct class, save one: the very event under discussion," which has a true likelihood of nothing less than 1 "and which we are trying to discover."⁸² The narrower reference class—the city block—certainly seems to be better evidence than the wider class—the entire road. But one can imagine a better reference class still, such as all buses on the city block within an hour of the accident. And if one had that data, there would yet be a better class, such as all buses on the city block within a *minute* of the accident. For all evidence of the kind offered in Blue Bus, there will be some better hypothetical reference class, the narrower one goes until one reaches the single event in question. It is easy to see that the Blue Bus evidence is flawed when we know that a narrower, and apparently more accurate, reference class would generate a

⁸¹ Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 109 (2007) [hereinafter *Problematic Value*].

⁸² *Id.* at 109.

different statistic. If the jury does not know about the narrower reference class, then something seems wrong with showing it the broader reference class.

While this criticism is apt in the Blue Bus hypothetical, not all statistical evidence cases suffer from the problem. Prison Yard, for instance, does not. The reference class in prison yard is a closed group: there are 25 prisoners in the yard, and all 25 are accounted for in the statistical evidence that the hypothetical posits. If one knew more information about narrower groups of prisoners within the 25, this would be better evidence to consider. But the fact that one does not have such information does not make the evidence irrelevant. The evidence in Prison Yard means just what it says it means. In this group of 25, 24 are guilty, and one is innocent. The evidence makes no special claim to being a “correct” base rate or a proper reference class for guilt, and a jury is unlikely to be confused by its meaning.

Similar examples featuring a closed population group may be fashioned to fit the preponderance standard. In the “Gatecrasher” hypothetical, originally developed by L. Jonathan Cohen, 1,000 people are in the bleachers at a rodeo, but it is known that only 499 people have bought tickets.⁸³ Thus, 501 of the people in the stands have not bought tickets.⁸⁴ Suppose the rodeo owner sues one of the 1,000 people who were in the stands, offering only this evidence. Like Prison Yard, the problem here is not that any reference class is being manipulated. It is stipulated that the reference class of 1,000 was chosen by the plaintiff because that is the number of people who were at the rodeo, and the statistical evidence’s meaning is clear. 499 people bought tickets, 501 people did not.⁸⁵

At its heart, the reference class problem has the potential to *confuse* jurors, and even trick them by hiding information from them. It is like Tribe’s anxieties about missing evidence. As with

⁸³ COHEN, *supra* note 12, at 49-120; L. Jonathan Cohen, *Subjective Probability and the Paradox of the Gatecrasher*, 1981 ARIZ. ST. L.J. 627, 627-29 (1981).

⁸⁴ David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101, 101 (1979).

⁸⁵ Another thing the Gatecrasher problem reveals is that the error costs of granting summary judgment to the defendant are *higher* than granting summary judgment to the plaintiff. See Brilmayer, *supra* note 58, at 676 (“Holding each rodeo spectator liable for trespass will result in 501 correct decisions and 499 incorrect decisions. Disallowing liability will result in only 499 correct decisions but 501 incorrect ones.”).

the pitfalls in the *Collins* case, jurors and judges can get confused in statistical evidence cases. The evidence in Blue Bus-style cases poses this risk. Litigants might choose among different reference classes (the county, the town, the road, the city block) to find the one class that best supports their claim, and present only that evidence to the jury. But why remedy that by banning statistical evidence altogether? Why not simply permit the opposing party to introduce its own statistical evidence, with a different reference class, in rebuttal?

Moreover, the evidence in Gatecrasher-style and Prison Yard-style cases does not seem to pose the same risk of reference class manipulation. And yet the intuitive judgment is the same: as a matter of law, the evidence is not enough to permit liability.

d. Public Acceptance

Finally, Charles Nesson has argued that statistical evidence verdicts are banned because they are unacceptable to the public.

Public acceptance “depends on a court’s ability to cast a verdict . . . as a statement about a past act—a statement about what happened.”⁸⁶ In statistical evidence cases, this is not possible. Statistical evidence “suggests a sufficiently high numerical probability of liability, but the absence of deference-inducing mechanisms in the judicial process is such that the public is unable to view a verdict against the defendant as a statement about what actually happened.”⁸⁷ Statistical evidence does not cause the public at large to accept the verdict, and because part of the trial process’s function is to generate verdicts acceptable to the public, judges are unwilling to let juries base their verdicts on statistical evidence alone. Thus, they give directed verdicts to defendants.⁸⁸ “What is crucial” about statistical evidence cases, Nesson notes, “is that the public cannot view whatever statement the factfinder makes as anything other than a bet based on the evidence.”⁸⁹

⁸⁶ See Nesson, *supra* note 7, at 1358.

⁸⁷ *Id.* at 1378.

⁸⁸ See *id.* at 1379. See also Kaye, *supra* note 20, at 40 (“[W]e would prefer not to advertise the fact that we are willing to sacrifice one innocent person in order to secure the conviction of nineteen guilty ones.”).

⁸⁹ See Nesson, *supra* note 7, at 1379.

Indeed, the public cannot. A verdict based on statistical evidence alone *is* a bet based on the evidence. For Nesson, it seems, the processes of trial evidence are directed at two goals: discovering the truth and generating verdicts that the public accepts. It is the public nature of courts and juries that bans statistical evidence verdicts.

Does the law permit verdicts generated purely by statistical evidence when the public will not find out the basis for the verdicts? This is unlikely. Imagine a different version of the Prison Yard hypothetical. Suppose that the yard was in a military facility, and the murder was not of a guard but of a government spy who had been sent to uncover information about a paramilitary organization, some of whose members reside in the prison. All evidence in the case is filed under seal, though the case remains governed by the reasonable doubt standard. It is hard to see how this would change our reading of the law. We do not become more comfortable with convicting the prisoner based on the statistical evidence alone when we know the public will not find out the basis for the verdict.⁹⁰ We still read the law as not allowing it. It is not the public nature of Prison Yard that is driving our view of the law.

The statistical evidence hypotheticals appear to prompt a *legal* judgment from readers. When confronted with Prison Yard, we might think that convicting the randomly selected prisoner is morally unjust. But we also conclude that it is *legally* impermissible. Our current law, as morally just or unjust as it may be, does not permit a person to be convicted based on statistical evidence alone. A judge's unwillingness to allow it is a function of the law itself. It is not an ad hoc maneuver to save the legal system in the eyes of the public by failing to properly apply the reasonable doubt standard.

In other words, if the public nature of courts is what is driving the legal interpretation of the standards of proof, a more basic theory of the law is required to explain how this is working. Why is

⁹⁰ One could argue that, even in such cases, the public *might* find out: jurors could divulge the evidence to the public after the trial, and that is what is driving the continuing intuition that the Prison Yard evidence is legally insufficient for conviction. Yet this remote anxiety does not seem to get at the essence of the concerns the hypothetical prompts.

the public nature of courts providing us with an interpretation of what the standards of proof require as a legal matter?

Still, in noting the connection with statistical evidence and betting, Nesson illuminated part of statistical evidence's true weakness. The public rejects verdicts based on statistical evidence because the public finds, correctly, that the standards of proof do not permit jurors to gamble. Nesson treats the public acceptability of verdicts as a question distinct from what the standards of proof ordinarily require as a legal matter. But perhaps the reason why statistical-evidence-generated verdicts are unacceptable is precisely that they fail to satisfy the standards of proof as a matter of law. The two considerations are one and the same, and the problem is juries making bets based on the evidence.

2. The Conjunction of Elements

To restate the conjunction paradox, plaintiffs and prosecutors must satisfy the relevant standard of proof with respect to *each* element, one-by-one.⁹¹ This permits, and, indeed, requires jurors to find defendants guilty when they do not have the required credence that the defendant committed the crime.

Consider the case of special verdicts, a not uncommon procedure in which a judge asks a jury to return separate verdicts on each element of a claim, offering a yes/no answer to each. Special verdict forms make conjunction paradoxes unavoidable.⁹² A negligence claim with three elements may return with a jury verdict saying each element has been met, though the jurors think the likelihood just barely clears 0.5 for each element. In so finding, the jurors will have been bound to rule that the defendant is liable, though they may not, and should not, have a credence above 0.5 that the defendant broke the law.

How might this puzzle be solved?

a. Rarity

One proposed solution has been to argue that conjunction paradoxes are rare in actual fact. Carl G. Wagner gives two reasons why this might be true.

⁹¹ See Cheng, *supra* note 7, at 1256; Schwartz & Sober, *supra* note 14, at 631.

⁹² Nesson, *supra* note 7, at 1387.

First, Wagner argues that conjunction paradoxes are rare because “defendants often dispute only one of the points in a conjunctive complaint, for example, conceding that a plaintiff indeed suffered injury, but denying the existence of proximate cause.”⁹³ This is an empirical argument, and perhaps not an implausible one. For strategic reasons, defendants might choose not to challenge certain elements of a plaintiff’s claim, in favor of presenting a cohesive, uncomplicated story to the jury.⁹⁴ Uncontested elements of a claim naturally may cause jurors to form credences very near 1 that these elements are satisfied, which minimizes the likelihood of a conjunction paradox.⁹⁵

Second, Wagner argues that when stated formally, the conjunction paradox frequently assumes that the elements’ likelihoods are independent, such that a 0.6 chance of A, a 0.6 chance of B, and a 0.6 chance of C means a 0.216 chance of ABC.⁹⁶ But in fact, elements’ likelihoods “are rarely, if ever, independent.”⁹⁷ Someone who is driving negligently is more likely to cause accidents than someone who is not, and thus the causation element, B, is not independent from the negligence element, A. If a person is driving negligently, a jury is justified in taking this fact as something that makes it more likely that this person *caused* the accident, absent factors to suggest otherwise.⁹⁸ Accidents, moreover, are likely to cause injuries, meaning the injury finding, C, is not unconnected to the causation finding, B. The negligence, causation, and injury elements thus aren’t independent. The 0.6 x 0.6 x 0.6 rule does not apply.

David S. Schwartz and Elliot Sober make a similar argument.⁹⁹ In the majority of cases, they argue, the “elements are probabilistically dependent, requiring the multiplication of *conditional* probabilities.”¹⁰⁰ These arguments are devoted to showing how conjunction paradox cases might occur infrequently in

⁹³ Wagner, *supra* note 15, at 1076.

⁹⁴ See *infra* Part II.A.

⁹⁵ See, e.g., Clermont, *supra* note 19, at 1108.

⁹⁶ See Wagner, *supra* note 15, at 1075.

⁹⁷ *Id.* at 1076.

⁹⁸ As Wagner puts it, “[w]ithout the independence assumption, we must employ the more general Pascalian multiplication rule.” *Id.*

⁹⁹ See Schwartz & Sober, *supra* note 14, at 655-58.

¹⁰⁰ *Id.* at 635.

real life—but they cannot eliminate them entirely.¹⁰¹ Whenever the elements of a claim are not completely *dependent* on one another,¹⁰² conjunction paradox cases will happen. They are just somewhat rarer than they would be if the elements were formally independent.

Indeed, one needs no grounding in formal probability theory to notice this. Suppose a plaintiff claims a defendant's negligent driving caused an accident. The defendant says that she was not driving negligently, and that, even if she had been, there was no way her negligent driving could have caused the accident. It is possible for a person to have been driving negligently, and to have become involved in an accident, yet still not have caused it.¹⁰³ Thus, a juror might come to some credence above 0.5 that each element—driving negligently and causation of the accident—is satisfied, but neither does nor should come to a credence above 0.5 that both elements are collectively satisfied.

Thus, even if conjunction cases are rare, the problem remains. All rarity arguments must acknowledge that conjunction paradox cases can happen. Would the defendants in such cases, having been duly convicted by a jury that did not think they had probably broken the law, be satisfied to learn cases like theirs are rare? They have a right not to be, particularly when cases like theirs could be prevented with relative ease. Along with abandoning special verdict forms, courts could instruct juries that the prosecutor or plaintiff must satisfy the burden of proof with respect to the whole claim, and not to each of its elements. To argue that the rarity of conjunction paradoxes solves the problem is to miss something about the problem's nature. It ignores, in Charles Nesson's

¹⁰¹ *Id.* at 656 (“To be sure, probabilistic dependence fails to eliminate the conjunction problem mathematically. . . . [T]he conjoint probability of elements is always less than the original individual probabilities.”).

¹⁰² By “completely dependent” I mean elements that are sufficient conditions for each other, of the form “Always, if *A*, then *B*; and always, if *B*, then *A*.” It is hard to imagine two separate elements of a claim being completely dependent in this sense, for this would seem to make the elements of the claim tautological.

¹⁰³ For example, an accident might have been caused by someone else's negligence—some third party's or the plaintiff's herself—rather than the defendant's. Or it might have been inevitable because of bad weather or something else.

observation, “the theoretical anomaly posed by the conjunction rule.”¹⁰⁴

b. A Judicial Mistake

Given that the conjunction paradox results from jury instructions and special verdict forms, some commentators have simply argued that these practices should be changed. If jury instructions were appropriately altered,¹⁰⁵ and special verdicts were disallowed,¹⁰⁶ the problem would disappear. And this is true. Had courts or legislatures enacted these changes, there would be no conjunction paradox. It is because they have not that the problem remains.

This is the paradox: if the standards of proof ask juries to make probability assessments, then why do courts tell juries to find that each element is satisfied rather than the whole claim? And why do courts sometimes give juries special verdict forms, which categorically require them to make a finding as to each element, and so to pronounce a defendant guilty, even when the jurors think—or should think—the overall likelihood of the claim being satisfied falls below the required threshold?

The reformers conclude that judges have made a mistake. Thus, implicit in the suggestion to change the law is the belief the law *does* require jurors to make probability assessments. It is the courts' failure to recognize how probability works that has caused them to write jury instructions in the way they have and to permit special verdicts in the way they do. This has led to an incoherence in how the law treats trial proof.

While this is certainly possible, it does not so much solve the paradox as give up on finding a solution. The contradiction regrettably remains, the reformers argue, because judges have not properly understood the law's probability requirements.

¹⁰⁴ Nesson, *supra* note 7, at 1387.

¹⁰⁵ See, e.g., Mark Spottswood, *Unraveling the Conjunction Paradox*, 15 L. PROB. & RISK 259, 294 (2016); Richard D. Friedman, *The Persistence of the Probabilistic Perspective*, 48 SETON HALL L. REV. 1589, 1594-95 (2018); Schwartz & Sober, *supra* note 14, at 690-91 (though they counsel that theorists should proceed cautiously when proposing legal reform); Hedden & Colyvan, *supra* note 58, at 452, 457-59; see also Nance, *supra* note 21, at 951-52 (arguing that jury instructions are frequently ambiguous and should be interpreted conjunctively where possible).

¹⁰⁶ See, e.g., Schwartz & Sober, *supra* note 14, at 690-91.

But it has been almost half a century since the paradox's discovery,¹⁰⁷ and courts show no signs of abolishing special verdicts or requiring conjunctive jury instructions. Perhaps they have not heeded the reformers' advice because the standards of proof require more from jurors than credences. This solution should at least be thoroughly considered before surrendering to the paradox.

c. Public Acceptance

Nesson finally argues that the public is more likely to accept verdicts if they are presented as "narrative history."¹⁰⁸ Because all narrative histories consist of many events, "[i]f we asked what the conjunctive probability of the narrative's independent elements is and dismissed the narrative when this probability was low, then we would have no history."¹⁰⁹ The trial process is thus "designed to produce a functional set of conclusions about what happened."¹¹⁰ Functional conclusions are acceptable ones, and "by refusing to adopt the conjunction rule," courts project their verdicts "as statements about what happened."¹¹¹ In so doing, the legal system forges, out of rough probabilistic material, pristine narrative histories to serve its purpose: deterring people from breaking the law.¹¹²

This argument provokes two objections. The first, a moral objection, is one that Nesson raises and does not resolve. Why should the truth be subjugated to the stories courts tell? It is not clear that Nesson believes this price is worth paying. "To argue that the search for truth may be compromised in order to enhance the power of the law's substantive message," he writes, "is to force us to

¹⁰⁷ See COHEN, *supra* note 12, at 66; Leubsdorf, *supra* note 4, at 1579 n.45 (tracing its origins to JEROME MICHAEL & MORTIMER J. ADLER, THE NATURE OF JUDICIAL PROOF: AN INQUIRY INTO THE LOGICAL, LEGAL, AND EMPIRICAL ASPECTS OF THE LAW OF EVIDENCE 141-42 (1931)).

¹⁰⁸ See Nesson, *supra* note 7, at 1388-90.

¹⁰⁹ See *id.* at 1389. A similar argument is made by Kevin Clermont, though he adopts the concept of "fuzzy logic" rather than narrative history. See Clermont, *supra* note 19, at 1089.

¹¹⁰ See Nesson, *supra* note 7, at 1389.

¹¹¹ *Id.* at 1390.

¹¹² *Id.*

confront an unsettling choice and to make an argument that is in some sense inherently unsatisfying.”¹¹³

It is especially unsatisfying when the message conflicts not only with the truth, but with the rights of the innocent to go free. Jurors in conjunction paradox cases are convicting defendants of crimes they do not—or, at least, should not—believe the defendant is likely enough to have committed. This means, if jurors’ credences track the truth, that juries in such cases are more likely to be convicting innocent defendants. This is a bargain that is hard to accept.¹¹⁴ A legal system that sacrifices innocent defendants for the publicity benefits of public narratives appears as a façade of justice, not as something to be defended precisely because it accepts that sacrifice.

The second objection is descriptive. Where in the *law* of the standards of proof does the limitation derived from public narratives come in? As a legal matter, the public acceptance solution to the conjunction paradox posits a complicated legal rule within the standards of proof; but the rule cannot be found anywhere in the text of the standards. Neither of the standards of proof state that judges should consider the publicity benefits to the justice system in deciding whether to allow certain cases to reach the jury. A more robust theory of the law and legal interpretation is thus going to be required if this view is to be sustained as something other than judges selectively departing from legal rules in order to impress the public.

Each of these solutions proposed by the credence account’s defenders thus fail to conclusively solve the problem.

II. OTHER ACCOUNTS OF THE STANDARDS OF PROOF

In light of the persistence of the statistical evidence and conjunction problems, scholars have developed alternative, non-credence-based accounts of the standards of proof, which have gained considerable ground in the academy. I briefly discuss three of the most promising solutions here: relative plausibility judgments, weight considerations, and probabilistic knowledge.

¹¹³ *Id.* at 1391.

¹¹⁴ *See id.*

A. Relative Plausibility Judgments

A new school of thought has argued against the credence account, suggesting that the standards of proof require jurors to reason in a wholly different way.¹¹⁵ On this argument, the jurors must make *relative plausibility* judgments with respect to the evidence at issue.¹¹⁶ Instead of telling jurors to convict only if they reach a certain probability assessment that the defendant broke the law, the standards are telling jurors to convict only after assessing the plausibility of each party's claims *in comparison with* what the other party claims happened.¹¹⁷

The process is straightforward enough. To determine the probative value of each item of evidence at trial, jurors must engage in a process of "inference to the best explanation."¹¹⁸ In a civil case, if there is a plaintiff and a defendant, each party will give their story of what happened. Under the preponderance standard, jurors must consider the evidence the plaintiff and defendant introduce, determining for themselves whose account the evidence supports. They must ask themselves, in other words, whose story is more plausible. At the end of the trial, having evaluated all the evidence, the jurors will have an answer. Whoever's account is more plausible wins.¹¹⁹ In a criminal case, the same reasoning applies, but it is skewed toward the defendant. Under the reasonable doubt standard, "the prosecution must provide a plausible account of guilt and show that there is no plausible account of innocence."¹²⁰ If the jurors can imagine *any* plausible account that a criminal defendant is innocent, then they are obliged to acquit the defendant.¹²¹

¹¹⁵ See, e.g., Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373 (1991); *Problematic Value*, *supra* note 81, at 107; Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 L. & PHIL. 223, 224-25 (2008); Ronald J. Allen & Alex Stein, *Evidence, Probability, and the Burden of Proof*, 55 ARIZ. L. REV. 557 (2013); Ronald J. Allen & Michael S. Pardo, *Relative Plausibility and Its Critics*, 23 INT'L J. EVID. & PROOF 5, 5-6 (2019); see also Allen & Leiter, *supra* note 21, at 1528.

¹¹⁶ See Allen & Stein, *supra* note 115, at 560.

¹¹⁷ *Id.* at 568.

¹¹⁸ See *Problematic Value*, *supra* note 81, at 136 (citation omitted).

¹¹⁹ See *id.* at 138 ("[T]he best explanation of the evidence concerning trials is that proof at trial involves inference to the best explanation from beginning to end.").

¹²⁰ See Allen & Leiter, *supra* note 21, at 1528.

¹²¹ *Id.* at 1531 n.125.

Relative plausibility theorists sometimes buttress their argument by pointing to empirical findings about how mock jurors behave, suggesting these findings are incommensurable with the credence account.¹²² The findings most frequently cited for this proposition are those of Nancy Pennington and Reid Hastie, who have argued that jurors undertake a “story model” of factfinding.¹²³ In their account, observations of mock jurors indicate that they assign “relevance to presented and inferred information” by deciding whether it fits into a cohesive narrative about the events in question, rather than weighing the evidence to come to a conscious subjective probability assessment.¹²⁴ Relative plausibility theorists thus criticize the credence account on the grounds that research has “made it rather plain that virtually no one thinks as the conventional legal theory requires.”¹²⁵

How jurors behave in practice remains a live question in the cognitive sciences.¹²⁶ But if jurors in real life do tend to reason along the lines suggested by the relative plausibility account, this does not mean that relative plausibility is a satisfactory account of what the standards require. This is because it is unclear how or why the legal meaning of the standards of proof should be influenced by juries’ behavior. Just as a jury might disregard the criminal statute it has in front of it,¹²⁷ or substitute its own lay understanding of a crime’s elements for what the judge has instructed,¹²⁸ jury

¹²² See, e.g., Pennington & Hastie, *supra* note 22, at 519.

¹²³ See *id.* at 520.

¹²⁴ See *id.* at 533, 545-46.

¹²⁵ Ronald J. Allen, *Factual Ambiguity and A Theory of Evidence*, 88 NW. U. L. REV. 604, 604 (1994) (citing Pennington & Hastie, *supra* note 22, at 519); see also Allen & Leiter, *supra* note 21, at 1527 (stating the relative plausibility approach was developed in response to the “empirical and analytical inadequacies of the expected utility and Bayesian approaches.”).

¹²⁶ For some proposed models of juror decision-making, see Valerie P. Hans & Valerie F. Reyna, *To Dollars from Sense: Qualitative to Quantitative Translation in Jury Damage Awards*, 8 J. EMPIRICAL LEGAL STUD. 120 (2011); Cass R. Sunstein, Daniel Kahneman, David Schkade & Ilana Ritov, *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153 (2002); Michael J. Saks & Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 LAW & SOC’Y REV. 123 (1980).

¹²⁷ See generally Alan W. Schefflin & Jon M. Van Dyke, *Merciful Juries: The Resilience of Jury Nullification*, 48 WASH. & LEE L. REV. 165 (1991).

¹²⁸ See, e.g., Dan M. Kahan, *Lay Perceptions of Justice vs. Criminal Law Doctrine: A False Dichotomy?*, 28 HOFSTRA L. REV. 793, 793 n.2 (2000) (citing Vicki L. Smith, *Prototypes in the Courtroom: Lay Representations of Legal Concepts*, 61 J. PERSONALITY & SOC. PSYCHOL. 857, 868 (1991)).

factfinding may deviate in practice from what the standards of proof require. In these cases, the jurors are simply failing to follow the law, not revealing the law's true nature. The relative plausibility account, like any account of the standards' *legal* requirements, must rise or fall on its conceptual merits.¹²⁹

Although some relative plausibility theorists cast the theory as a repudiation of quantitative methods,¹³⁰ Professor Edward K. Cheng offers a way to synthesize it with Bayesianism. If the preponderance standard is a *likelihood ratio*, and jurors are engaging in inference to the best explanation when they evaluate evidence, then Bayesianism and relative plausibility can be reconciled.¹³¹ Inference to the best explanation still means inference to the best explanation *offered by either party*, for each side must offer the jury its own account of what happened.¹³² If the likelihood ratio of each side's story in the juror's mind exceeds 1:1 in favor of the plaintiff's story, then the plaintiff wins. If the ratio equals or exceeds 1:1 in favor of the defendant's story, the defendant wins.¹³³

This offers a way to solve the statistical evidence and conjunction problems. The problem with the evidence in *Blue Bus* is the improper *connection* between it and the plaintiff's story of what happened. "It is hard to envision how the identity of the bus on the night of the accident, without more, gives us much, if any, information on the proportion of blue buses owned by the defendant."¹³⁴ And relative plausibility solves the conjunction problem. When a juror assesses the likelihood ratio to be more than

¹²⁹ For example, see how L. Jonathan Cohen described the inquiry into what the standards of proof legally require. L. Jonathan Cohen, *The Role of Evidential Weight in Criminal Proof*, 66 B.U. L. REV. 635, 635 (1986) ("And the thesis for which I have thus contended is essentially a normative one, concerned with answering the question 'What is the legally correct way to judge proofs?,' not a factual one, concerned with answering the question 'What is the way in which proofs are actually judged?').

¹³⁰ See *Problematic Value*, *supra* note 81, at 136 (stating that the relative plausibility account "does not much depend on the quantification of the value of individual items of evidence . . .").

¹³¹ See Cheng, *supra* note 7, at 1278.

¹³² See *Problematic Value*, *supra* note 81, at 136.

¹³³ Similarly, the reasonable doubt standard may be cast as a threshold requirement for a higher likelihood ratio. See Cheng, *supra* note 7, at 1278.

¹³⁴ *Id.* at 1270. A similar observation—involving counterfactuals—is made by Lara Buchak. See Buchak, *supra* note 26, at 294-96. See also *infra* Part III.

1:1 with respect to each element of a claim, she is logically guaranteed to end up with an assessment of the likelihood ratio of all the elements that is greater than 1:1.¹³⁵

But as an explanation of what the standards of proof require as a legal matter, relative plausibility leaves unsatisfactory gaps.¹³⁶ It may be true that persuasive advocacy often relies on telling a compelling and coherent story to the jury, and that juries expect the defendant to tell them a coherent story about what happened, rather than simply to attack the plaintiff's story. But it is another thing altogether to conclude that the jurors are *forbidden by law* from drawing their own conclusions about what happened in the case.

Imagine a case in which the defendant is accused of civil theft.¹³⁷ There are two witnesses, A and B, each of whom say they saw the defendant steal an emerald from a closed jewelry store one night. The plaintiff jeweler tells the jury a story—a simple one, consistent with the witnesses' testimony: the defendant stole the jewel. The defendant tells a very different story. The jeweler is collaborating with the witnesses, as the three of them took the jewel together as part of an insurance scheme, sold it, and are now framing the defendant to escape liability.

Imagine a juror, after reviewing the evidence, concludes that neither story is true. In fact, it is Witness A, alone, who is framing the defendant. Witness B is merely confused, as she is shortsighted, and mistakenly thought she saw the defendant pocket the emerald, when she actually saw Witness A in a disguise. The jeweler, meanwhile, honestly believes the witnesses' accounts.

Let us say that the juror's credence in the plaintiff's story is now 0.0003, and her credence in the defendant's story is 0.0002. She

¹³⁵ See Cheng, *supra* note 7, at 1263-65.

¹³⁶ Relative plausibility theorists have sometimes cast their arguments as aiming at broader conclusions than interpretations of the standards of proof. See, e.g., *Relative Plausibility*, *supra* note 115, at 7 (“We focus on burdens of proof . . . as a lens through which to observe the legal system, but [the] burdens of proof are only a lens. What is being observed is the entire litigation process, which includes burdens of proof as one crucial component.”).

¹³⁷ Thefts are generally civilly actionable under the tort of conversion. See, e.g., *Cnty. Bank v. Courtney*, 884 So. 2d 767, 774 (Miss. 2004) (quoting *First Investors Corp. v. Rayner*, 738 So. 2d 228, 234-35 (Miss. 1999)) (describing conversion as “an ‘intent to exercise dominion or control over goods which is inconsistent with the true owner’s right.’”).

has not entirely ruled out either of them, and she thinks the plaintiff's straightforward explanation is perhaps slightly more plausible than the defendant's conspiracy theory. But she thinks both are extremely unlikely. Her credence in her own conclusion, on the other hand—that it is Witness A alone who is framing the defendant—is high, at least 0.9.

If this juror simply compared the plausibility of the plaintiff's and defendant's accounts, she would have to choose the plaintiff's. Yet this juror is clearly allowed to find the defendant not liable. She does not think the defendant took the emerald. But in finding the defendant not liable, she would have chosen neither story. She would not have compared the plausibility of the two competing accounts, but rather she would have drawn her own conclusion, separate and apart from both.

The relative plausibility account seems bound to insist that she cannot do this. Relative plausibility is only distinguishable from the credence account when the law limits the defendant to a single story, or perhaps to a short menu of possible stories. Otherwise, the likelihood ratio—expressed in non-numerical terms, the relative plausibility judgment—would simply be a credence. The comparison would not be between the plaintiff's and defendant's story of the case, but between the plaintiff's story and every other way the world could be interpreted. Evaluating the latter comparison is no different from forming a credence as to whether the defendant broke the law. For if a juror can draw her own conclusion, the relative plausibility judgment is now, again, just a credence. In Bayesian terms, if the likelihood ratio of the plaintiff's story to the other possible ways the world could be is 1:3, it is a credence of 0.33 that the plaintiff's story is true.

Can the relative plausibility account be rescued? One possible thought is that the juror was never permitted to draw her own conclusion in the first place. She was not allowed to determine in her own mind that it was Witness A who was framing the defendant. But how could she not be so permitted? She has been told by the judge to evaluate the evidence in order to see whether the defendant has broken the law. She has followed this directive. She has determined that the defendant has not broken the law, because he is being framed by Witness A. This is the jury's role.

Another possible thought is that the juror in the emerald case has permissibly substituted the defendant's provided narrative for a better narrative in the defendant's favor. As Cheng notes, "if the defendant fails to provide a narrative, the jury will simply substitute the best narrative it can construct in favor of the defendant."¹³⁸ But more needs to be said about how this juror's reasoning could differ from the credence account. By choosing another narrative, one unrepresented by the defense, to compare with the plaintiff's, the juror has evaluated alternative possibilities. And if she is not limited to the possibilities presented by the parties, and she has arrived at a likelihood ratio that the defendant broke the law, then she has simply estimated the probability that the defendant broke the law.

Thus, the relative plausibility account confronts a choice. Either it insists that the law mandates that jurors find civil defendants liable in cases when they think those defendants did not actually break the law, or it collapses into the credence account. Choosing the former might resolve the statistical evidence and conjunction problems, but it seems impossible to sustain as a measure of what the preponderance standard requires. Choosing the latter leaves us back at square one: the credence account. The relative plausibility account may be capable of resolving cases like these, but until then, it remains an incomplete account of what, as a legal matter, the standards of proof ask of jurors.

B. Weight

Another thought has been to resolve the statistical evidence problem by distinguishing between evidence's relevance and its *weight*. The archetypal statement of how these two qualities differ was offered by John Maynard Keynes.¹³⁹ As the amount of relevant evidence available to us goes up, he suggested, the probability that a claim is true will increase or decrease.¹⁴⁰ It will increase if the new evidence supports the claim, and it will decrease if the new evidence undermines the claim.¹⁴¹ But in either circumstance, some

¹³⁸ Cheng, *supra* note 7, at 1262 n.15.

¹³⁹ See generally JOHN MAYNARD KEYNES, A TREATISE ON PROBABILITY (1921).

¹⁴⁰ *Id.* at 71.

¹⁴¹ *Id.*

other quality in the evidence will always have gone up: the weight of the evidence on which we are basing our judgment.¹⁴² “[S]omething seems to have increased in either case,—we have a more substantial basis upon which to rest our conclusion.”¹⁴³ New relevant evidence, whether incriminating or exculpating, “increases the *weight* of an argument. New evidence will sometimes decrease the probability of an argument, but it will always increase its ‘weight.’”¹⁴⁴

On this idea, evidence has two qualities, both of which are relevant to how it should be evaluated. Someone might look out the window one morning and conclude, based on the cloudy sky, that there is a 60% chance it will rain that afternoon. Later, that same person might look at the weather report online and discover that it also suggests there is a 60% chance of rain that afternoon. The new evidence from the weather website should not change her credence that it will rain: it was 0.6 after she looked out the window, and it remained 0.6 after she checked the website. But something else has changed: she has *more* evidence on which to base her credence. The evidence has accumulated more weight.

The problem, on this view, with statistical evidence is not the insufficiency of its probativeness, but the lightness of its weight. The statistical evidence hypotheticals thus reveal an additional requirement of the standards of proof. Prison Yard reveals that the reasonable doubt standard requires the evidence of guilt to be of “reasonable completeness.”¹⁴⁵ The credence account may be right that the jury is required to form a credence of at least 0.95 that the defendant is guilty, but this is not all the jury is required to do. The jury must have evidence of enough weight to base its decision, and the problem with Prison Yard is that the evidence at the jury’s disposal is too light. Like several of the solutions proposed by the credence account’s defenders, this argument explains why a juror might find statistical evidence insufficient for guilt. To justify *judges’* granting summary judgment to defendants in statistical evidence cases before they reach a jury at all requires one further step.

¹⁴² *Id.*

¹⁴³ *Id.* See also Hock Lai Lo, *supra* note 9, at § 3.3.

¹⁴⁴ KEYNES, *supra* note 139, at 71.

¹⁴⁵ See Cohen, *supra* note 129, at 649.

Dale Nance, a contemporary scholar of Keynesian weight, takes this step, arguing that the question of how much weight is enough to prompt a decision of guilt or liability cannot be answered by a jury because it is a “question of law.”¹⁴⁶ Thus, the statistical evidence problem is solved using Keynes’s division. The judge’s role is to ensure that evidence carries enough weight to satisfy the legal standard of proof; the jury’s role is to evaluate the evidence’s probativeness. Courts must grant summary judgment to defendants in statistical evidence cases because the evidence is too light to license verdicts of guilt or liability as a matter of law.

This explanation illuminates the problem. There is something wrong with statistical evidence, and the postulation of Keynesian weight is an elegant statement of what’s wrong with it. Statistical evidence is insufficiently weighty. But what is it *about* statistical evidence that makes it categorically lightweight as a matter of law? Why are judges in the business of determining the threshold weight of trial evidence, and why doesn’t statistical evidence exceed that threshold? That is the question.

One answer is that statistical evidence has an improper causal relationship with the claim to be proved.¹⁴⁷ L. Jonathan Cohen made a similar argument, concluding that statistical evidence was insufficiently weighty because it failed to license a particular style of factfinding.¹⁴⁸ Factfinders must discover the causes of things, and there is no causal relationship between statistical evidence and the claim it seeks to prove. Imagine, he suggested, we are asking whether a teenager is likely to reach the age of 70. We learn that the teenager’s last name contains 6 letters, and it turns out that the frequency of people living past 70 who have 6-letter last names “is greater than the frequency in the population at large.”¹⁴⁹ This difference in frequencies does not make it any more likely that this young person will live past 70. Some causal story is required to

¹⁴⁶ DALE A. NANCE, *THE BURDENS OF PROOF: DISCRIMINATORY POWER, WEIGHT OF EVIDENCE, AND TENACITY OF BELIEF* x-xi (2016).

¹⁴⁷ These arguments share a similarity with those of Judith Jarvis Thomson, who argued that the problem with statistical evidence is that it does not have the right sort of causal relationship with the claim to be proved. *See* Thomson, *supra* note 28, at 202-03.

¹⁴⁸ *See, e.g.*, COHEN, *supra* note 12, at 116-20.

¹⁴⁹ L. Jonathan Cohen, *Subjective Probability and the Paradox of the Gatecrasher*, 1981 ARIZ. ST. L.J. 627, 633 (1981).

show why having a 6-letter last name makes the person more likely to become a septuagenarian. Otherwise, the statistical evidence is mere coincidence.

Thus, in Gatecrasher, there is no “*causal link*” between the statistical evidence and the fact that a randomly selected rodeo attendee failed to buy a ticket.¹⁵⁰ The evidence is mere coincidence.

Yet the case of the septuagenarian is different from Gatecrasher. The evidence in Gatecrasher is that 501 out of the rodeo’s 1,000 total attendees were trespassers. The randomly selected attendee was certainly present at the rodeo, and so he is *part* of the relevant group of people who are slightly more likely to have trespassed than not. The appropriate analogy is to imagine you are being asked to determine whether a stranger whose name you do not know has a 6-letter last name. Suppose you learn that the stranger is over 70, and that 51% of people who are currently over 70 have 6-letter last names. There is still no causal story to explain why this is the case, and yet now it is not unreasonable to take this statistic as evidence that the stranger has a 6-letter last name.

A second problem with the argument about the lack of a proper causal relationship is that it seems to lead inexorably to the conclusion that statistical evidence is not *relevant* to legal claims, and so should be excluded wholesale from jury trials. But it is not. As long as other evidence is available in the case, and as long as it is not confusing, then courts do trust juries, perhaps with appropriate limiting instructions, to evaluate it.¹⁵¹

The puzzle of statistical evidence is that it appears to be different—when offered on its own, with no other evidence available—from the sorts of complicated technical and scientific evidence that courts permit juries to see on a regular basis. This leaves the question of why judges must grant summary judgment to the defendant in statistical evidence cases. Behind the argument from Keynesian weight, Tribe’s anxieties about jury confusion echo. But the legal rule seems to be different when *only* statistical evidence is available in a case, and the question remains why.

¹⁵⁰ See *id.* (emphasis added).

¹⁵¹ See Koehler, *supra* note 60, at 377-80 (collecting cases). A salient example is DNA evidence, which frequently appears in cases in conjunction with other kinds of evidence establishing, for example, motive or a connection between a defendant and the crime.

C. Probabilistic Knowledge

Most recently, the philosopher Sarah Moss has developed a groundbreaking argument in epistemology and the philosophy of language.¹⁵² On this argument, people have probabilistic beliefs, and these beliefs come in different kinds.¹⁵³ A person who is more confident than not that it will rain this afternoon might have a credence of 0.6 that it will rain this afternoon. But she also might believe outright that it is more likely than not that it will rain this afternoon, and this belief may constitute knowledge. She might know that it is more likely than not that it will rain this afternoon, and this is different from merely believing it will rain this afternoon.

According to Moss, the preponderance standard of proof requires the factfinder to know that the defendant probably broke the law.¹⁵⁴ “A defendant is proved liable by a preponderance of the evidence only if the factfinder has greater than .5 credence that the defendant is liable, *and that probabilistic belief constitutes knowledge.*”¹⁵⁵ Similarly, the reasonable doubt standard requires the juror to know a separate “probabilistic content, namely that the likelihood that the defendant is guilty exceeds a certain threshold.”¹⁵⁶ The likelihood is somewhere between 0.5 and 1—perhaps 0.9, or one of the other probabilities frequently cited by the credence account as sufficient for reasonable doubt.¹⁵⁷ On this view, the traditional credence account is correct about the level of the subjective probability assessment required—0.5 for the civil standard, somewhere close to but below 1 for the criminal standard¹⁵⁸—but something else is also required. The jurors cannot merely believe that these likelihoods are satisfied. They must know it.

This, Moss argues, solves the statistical evidence problem. If you are a juror in the Prison Yard case, “you may believe that the

¹⁵² See generally MOSS, *supra* note 10.

¹⁵³ *Id.* at 3-4.

¹⁵⁴ *Id.* at 210.

¹⁵⁵ *Id.*

¹⁵⁶ *Id.* at 212.

¹⁵⁷ *Id.*

¹⁵⁸ Moss notes that the probabilistic knowledge solution also works when the criminal standard is defined as probability 1. See *id.*

defendant is .96 likely” to have committed the crime, but you cannot know it.¹⁵⁹ Specifically, “the prosecution cannot provide you with knowledge of this content merely by proving that 24 of the 25 prisoners in the yard are guilty, since there is a certain possibility that you cannot rule out—namely, that the defendant is less likely to be guilty than an arbitrary prisoner in the yard.”¹⁶⁰ And the same is true of Gatecrasher, in which “the plaintiff cannot provide you with knowledge” that the rodeo attendee probably trespassed, because you cannot rule out the possibility “that the defendant is less likely to have climbed over the fence than an arbitrary person at the rodeo.”¹⁶¹ By contrast, if an eyewitness whom you know is probably telling the truth testifies that she saw the defendant jump over the ticket booth, “you can thereby come to know the probabilistic content that the defendant” probably trespassed, and so find him liable.¹⁶²

Let us consider more closely whether probabilistic knowledge solves the problem. Imagine a jar that contains 96 blue marbles and 4 red ones. You draw a marble but do not look at it, keeping it in the palm of your hand. Before drawing the marble, you examined the jar carefully, concluding it was just like normal jars, and you counted the marbles inside it, noting their colors. It is plausible now to say that you *know* there is a 0.96 likelihood the marble you have drawn is blue. Indeed, this seems like a paradigm case of probabilistic knowledge.

It is also the same attitude that a juror in Prison Yard might have. If the juror assesses the witness who saw the prisoners at a distance, and judges the witness to be truthful, then she has examined the source of the 0.96 evidence in the same way that you have examined the jar with 96 blue marbles in it. The juror seems entitled to say she knows the defendant probably committed the murder in just the same way as you are entitled to say you know the marble in your hand is probably blue. The statistical evidence problem remains unresolved, at least without further parameters to sustain the theory.

¹⁵⁹ *Id.* at 213.

¹⁶⁰ *Id.*

¹⁶¹ *Id.*

¹⁶² *Id.*

The parameter Moss offers is that probabilistic knowledge about *people* is different from probabilistic knowledge about marbles in a jar. On this argument, knowledge is contingent on one's ability to rule out relevant alternative possibilities.¹⁶³ The reason a juror cannot acquire probabilistic knowledge in Prison Yard is that there are relevant possibilities the juror cannot rule out. "You do not know," Moss writes, "that this particular defendant is probably liable, because you can't rule out a relevant possibility that is inconsistent with this content—namely, that the defendant is an individual whose character makes him far less likely to" have committed the crime than an arbitrarily selected prisoner in the yard.¹⁶⁴

It is the complexity of people's *characters* that makes deriving probabilistic knowledge about them from statistical evidence impossible. Marbles are not entitled to insist that they are individuals whose makeup makes them less likely to be blue than the other marbles in the jar. People are, because people have unique characters. This explains why deriving probabilistic knowledge about people from the statistical evidence is impossible.

The problem with this argument is that it doesn't distinguish between statistical evidence cases and non-statistical evidence cases. People can make an objection from individual character when faced with many kinds of evidence, not just the statistical kind. If a person's individual character might raise relevant alternative possibilities in statistical evidence cases, then it might do so in almost all cases.

Suppose, for example, the Prison Yard hypothetical is changed to remove its statistical aspect. Now, the witness in the case was closer to the scene, and she tells the jury that she saw *this very defendant* take part in the murder. The jury is also told that a witness's perceptions from this distance are 96% accurate. This is now a case, most lawyers think, in which a jury could find the defendant guilty beyond a reasonable doubt. It is also, seemingly, a

¹⁶³ Sarah Moss, *Knowledge and Legal Proof*, in 7 OXFORD STUDIES IN EPISTEMOLOGY 23 (forthcoming 2021).

¹⁶⁴ *Id.* at 24.

case in which the jury could *know* that the defendant probably broke the law.¹⁶⁵

But this defendant can make the same objection from individual character. He can say he is an individual whose character makes him far less likely to have committed the crime than any arbitrarily selected person whom the witness might claim to have seen, and that because of this, the witness is much more likely to have misidentified *him* than it would appear from the witness's general 96% accuracy rate. The two Prison Yard cases are the same in this respect, and so, if this is the solution to the statistical evidence problem, then the outcome should be the same in both. In neither case can the jurors rule out the relevant alternative possibility posed by the prisoner's individual character. They cannot acquire probabilistic knowledge as to his guilt.

And if the objection from individual character can be raised for forms of evidence that *can* license guilt as a matter of law, then probabilistic knowledge no longer solves the statistical evidence problem. A person's individual character is just as unconnected from an eyewitness's general accuracy rate as it is from the statistical evidence.

III. THE OUTRIGHT BELIEF REQUIREMENT

Suppose the standards of proof, civil and criminal alike, require jurors to convict and hold liable defendants only if they come to an outright belief—a credence of 1—that the defendant broke the law. The statistical evidence and conjunction problems vanish.

There is no conjunction paradox if a credence of 1 is required for a given standard of proof. For reasonable jurors to have a credence of 1 in the conjunction of the elements, they must always have the same credence in each element: 1. Likewise, statistical evidence is insufficient for liability because it cannot justifiably induce an outright belief that the defendant broke the law. At best, Blue Bus induces a credence of 0.8 that the defendant caused the

¹⁶⁵ This is also the kind of case in which laypeople appear to be more comfortable with coming to a guilty verdict as mock jurors. *See generally* Wells, *supra* note 64.

plaintiff's injury.¹⁶⁶ At best, Prison Yard induces a credence of 0.96 that the defendant took part in the murder.

Lara Buchak gives an epistemological explanation for why evidence of the kind found in Blue Bus cannot reasonably give rise to an outright belief.¹⁶⁷ Counterfactual reasoning reveals the lack of a causal connection between the evidence and the claim to be proved.¹⁶⁸ If the plaintiff in Blue Bus had produced as evidence scraps of blue paint she found on her car after the accident, which matched the color of Blue Bus buses and only Blue Bus buses, a jury might reason in the following way. "If the bus at issue had not belonged to the Blue Bus company, then it would not have left this color of paint at the scene." This is a plausible claim. But a jury deliberating over the 80% market share evidence alone could not construct a plausible counterfactual of this kind. "If the Blue Bus company had not caused the accident, then it would not have operated 80 percent of the buses on this particular road" seems to make less sense, if it makes any sense at all.¹⁶⁹

But how can an outright belief requirement in the standards of proof be sustained? How often can anything be proven to a full certainty, which is to say, beyond any doubt? All the criminal standard asks for is proof beyond a *reasonable* doubt. And how can both standards require an outright belief without collapsing into one another? How can the criminal standard be harder to satisfy than the civil standard, if both standards require the jurors to form an outright belief that the defendant broke the law?

These questions can be answered easily enough if the standards of proof take a certain view of how people think through the things they learn about the world. People form outright beliefs about things, not just credences. And their willingness to form outright beliefs depends, in part, on the circumstances in which they find themselves when they are making up their minds.

¹⁶⁶ See Buchak, *supra* note 26, at 291.

¹⁶⁷ *Id.* at 292-93 (noting some epistemological literature has claimed "purely statistical evidence should not produce belief").

¹⁶⁸ See *id.* at 294.

¹⁶⁹ See *id.* (considering a hypothetical non-legal case involving theft).

A. People Have Outright Beliefs

To fix ideas about the distinction between an outright belief and a credence of less than 1, consider two people walking in a forest. Both hear a persistent tapping noise on a nearby tree, hidden from view. Both are familiar with the woods, and with what a woodpecker's drumming sounds like. The first person decides there is a 0.95 chance that this noise is caused by a woodpecker. She is very confident, but she has considered the odds that she may be wrong, and has concluded that there is a 0.05 chance, which she cannot rule out, that it is not a woodpecker, and that it is a man-made noise or something else.

By contrast, the second person believes outright that a woodpecker is making the noise. Where the first person has consciously estimated a high likelihood that it is a woodpecker, the second person isn't thinking about likelihood or chance at all, and, if prompted to do so, would say the chances are 1, or 100% that it is a woodpecker. The first person has a credence of 0.95; the second person has a credence of 1, or an outright belief.

One might object in the following way. Does the second person really have no doubts at all that a woodpecker is making the noise? She has not seen the woodpecker. And even if she had seen it, there would be some chance, however remote, that it was an illusion—or that something was temporarily wrong with her eyes. Indeed, when pressed, it is hard to imagine evidence of any kind that cannot be mistaken or faked.¹⁷⁰ Thus, one might conclude that outright beliefs do not exist. This hiker might be *very* confident it is a woodpecker, but is still leaving some room, even if small, for doubt.

Still, on reflection, outright beliefs seem to be commonplace. For example, they form the basis of our credences. When we read the weather report and conclude there is a 0.6 chance of rain this afternoon, we have in our minds a credence. But on what grounds have we made this assessment? There are possibilities that we are

¹⁷⁰ This concern is evident for courts. One judge has argued that a standard of proof: [C]an never be set at certainty or 100% probability, because [t]ime is irreversible, events unique, and any reconstruction of the past at best an approximation. As a result of this lack of certainty about what happened, it is inescapable that the trier's conclusions be based on probabilities.

United States v. Fatico, 458 F. Supp. 388, 411 (E.D.N.Y. 1978) (citing MAGUIRE, ET AL., CASES AND MATERIALS ON EVIDENCE 1 (6th ed. 1973)).

taking for granted—that we are taking as true. For example, why would we have ended up with a credence of 0.6 that it will rain unless we're certain—credence 1—that the meteorologist actually reported a 60% chance of rain? We are assuming that we did not mishear, or misremember, the content of the weather report. We have a credence of 1 that we remember the weather report correctly. If we did not, then we should have factored that into our assessment of the rain's likelihood, and so ended up with a credence somewhere above or below 0.6 that it will rain. This we did not do. Our credence remains 0.6.

Of course, if so prompted, we might start doubting our certainty that we remember the meteorologist reported a 60% chance of rain. We might thus try to determine what our credence that we correctly remember the weather report really is. In trying to determine this credence, we would find still more assumptions. We might ask ourselves how frequently we misremember things, taking this as evidence for whether we have remembered the weather report correctly. But this assessment requires basic assumptions, including the idea that the present is *like* the past—that we have not developed some sudden memory deficiency, or that the world has not radically changed in the last few moments. Eventually, even the most sophisticated observers would run out of time to assess these background propositions as credences below 1.

Outright beliefs, then, appear to be necessary for us to go about our lives.¹⁷¹ Ideal Bayesian reasoning “isn't feasible for cognitively limited agents like us, and so we need an attitude of *outright belief* or of *settling on the truth of propositions*, so as to limit what we consider in our reasoning to possibilities consistent with what we have settled on.”¹⁷² The world is vast and our cognitive capacities are limited. To abandon outright beliefs entirely is to “risk being overwhelmed by the huge mass of uncertainty that the approach generates.”¹⁷³ There is such a thing as outright beliefs, which people adopt and as to which they do not retain any doubts. Of course,

¹⁷¹ See, e.g., Greco, *supra* note 24; Ross & Schroeder, *supra* note 25, at 286.

¹⁷² Ross & Schroeder, *supra* note 25, at 286.

¹⁷³ Richard Holton, *Intention as a Model for Belief*, in RATIONAL AND SOCIAL AGENCY: THE PHILOSOPHY OF MICHAEL BRATMAN 12, 14 (Manuel Vargas & Gideon Yaffe eds., Oxford Univ. Press 2014).

subsequent things might happen to destabilize the outright belief, giving us doubts, as the next section will suggest.

This in turn may offer another reason why it is *unreasonable* to form an outright belief that the defendant broke the law in the statistical evidence cases. Outright beliefs are useful because they are necessary mental shortcuts in an uncertain and complicated world. But it takes no mental work at all to recognize that statistical evidence below 1 supports a probability less than 1. At best, the odds in Prison Yard are 0.96 that the defendant broke the law. To take the mental shortcut to a credence of 1 based on the evidence saves no time or resources, while it exacerbates the chances of making an incorrect judgment. It is an unreasonable thing to do.

B. People's Willingness to Form Outright Beliefs Depends on the Circumstances in Which They Are Making Up Their Minds

If people do have outright beliefs, it is not hard to see how the circumstances of their decision-making might influence their willingness to form such beliefs. Consider one more thought experiment.

Imagine a stranger asks you what time it is.¹⁷⁴ You look at your watch and answer, "It's 9:55 a.m." He asks if you are sure that it's 9:55 a.m. You answer, "I'm sure. I have looked at my watch, and my watch is correct." This is not an unreasonable answer, assuming you take some care to check your watch is timed correctly. In other words, it does not seem unreasonable for you to have an outright belief that it is 9:55 a.m.

But now imagine the stranger offers you a bet. If it really is 9:55 a.m., he will give you a dollar. But if you are wrong, and it isn't 9:55 a.m., then you must give him everything you have—all of your worldly possessions.

An intuition quickly emerges that you should not accept this bet. But why not? To be sure, the potential reward is low, and the potential loss is very high, but if you really evaluate the likelihood that it is 9:55 a.m. to be 100%, then you are bound to win. You are guaranteed to be a dollar richer than you were before. The fact that

¹⁷⁴ This example, as well as the conceptual analysis that follows, is derived from Greco, *supra* note 24, at 186.

you are unwilling to make the bet suggests that you are not fully confident that it is 9:55 a.m., and that you have some credence near 1, but below it—specifically, below the ratio of how much you value everything you have to how much you value one dollar.¹⁷⁵

What does this reveal? Perhaps you were wrong in saying you were 100% sure it was 9:55 a.m. when the stranger first asked you. You actually had a credence somewhere below 1 that it was 9:55 a.m., but you mistakenly thought you had a credence of 1. This possibility is what might drive an argument that there are no such things as outright beliefs. But for the reasons offered in the previous section, this view seems wrong. There are outright beliefs, because people need them in order to function, and it is not unreasonable to say that when many people look at their watches, they form an outright belief that the time is what the watch says.

A more plausible possibility is that you really had a credence of 1 that it was 9:55 a.m. when you were asked the first time, but that something *changed* when you were offered the high-stakes bet. Two things, in particular, might have changed. For one thing, you have new evidence, heretofore unavailable, about whether it really is 9:55 a.m. The fact that a stranger has just offered you this bet might indicate that he knows something you don't. Perhaps he accessed your watch earlier that morning and changed it, while you were unaware, and you did not consider this possibility before, but now that you have been offered the bet, it has entered your mind. Considering it has destabilized your outright belief that it is 9:55 a.m. into a credence somewhere below 1.

This is certainly possible. But something else has changed too. The consequences to whether it is 9:55 a.m. have changed. Comparatively little depended on your outright belief that it was 9:55 a.m. when you were first asked, but after being offered the bet, very much depended on it. This radical change in circumstances is what caused you to reassess your outright belief that it was 9:55 a.m., and, as a result, you now have a credence somewhere below 1

¹⁷⁵ Another possibility is that the intuition telling you not to take the bet is an irrational one. We might be used to using heuristics when making decisions, and one plausible heuristic is “don't ever take bets that are so one-sided.” This heuristic in this case is leading us astray. You are failing to act in such a way as to guarantee a profit of one dollar. And yet it still does not *seem* at all irrational to decline the bet, even after prolonged reflection.

that it is 9:55 a.m. As the philosopher Daniel Greco has argued, the “possibilities one takes seriously—which possibilities one treats as ‘live’—is sensitive to a wide range of situational factors, including practical ones.”¹⁷⁶ Stated differently, the things people believe outright are sensitive to the consequences of those beliefs.¹⁷⁷

This answers the second objection to the outright belief requirement. Both standards of proof can require the jurors to form an outright belief that the defendant broke the law, and the criminal standard can still be harder to satisfy than the civil one. Jurors in the circumstances of a civil case might be more willing to form an outright belief that the defendant broke the law than jurors in the circumstances of a criminal case. And *that* is what the standards of proof are communicating. They are telling the jury about the circumstances in which the jurors have found themselves, and about the law’s view of the consequences that flow from forming an outright belief that the defendant broke the law. The circumstances of a criminal case are, in the law’s judgment, different from a civil one.¹⁷⁸

Older judicial language offers support for this idea. One court has asserted that the preponderance of the evidence standard is satisfied if the tribunal arrives at an “actual belief in its truth, derived from the evidence . . . notwithstanding any doubts that may still linger there.”¹⁷⁹ Another has suggested:

The burden of proof requires the party carrying it to prove to the jury the facts, upon which his case or affirmative defense depends, by a preponderance or greater weight of the credible evidence. This means merely that the party, who has the burden of proof, must produce evidence, tending to show the

¹⁷⁶ Greco, *supra* note 24, at 186.

¹⁷⁷ On this view, even minor changes in circumstance can destabilize outright beliefs. For example, if someone asks you the time, and you do not have a watch, but you see a public clock some distance away that says 9:55, you might draw an outright belief that it is 9:55 and say so. But if the person then asks, “Are you *sure*?” this might change the circumstances of your belief-formation and might in itself cause you to revisit the outright belief. You might reply with something along the lines of “I guess I’m not sure, because that clock could be wrong.”

¹⁷⁸ That is one reason why the standards of proof prohibit an attorney from trying to change the circumstances of the case from what the standards say they are. A civil defense attorney could not, for example, in her closing statement offer the jurors a high-stakes bet on whether the defendant actually broke the law.

¹⁷⁹ *Sargent v. Mass. Accident Co.*, 29 N.E.2d 825, 827 (Mass. 1940).

truth of those facts, “which is more convincing to them as worthy of belief than that which is offered in opposition thereto.”¹⁸⁰

It is hard to say for sure what courts mean in such statements. There is a fluidity in the everyday words we use to describe different states of belief and credence.¹⁸¹ But one thing they might mean is that the preponderance standard is asking jurors to form an outright belief that the defendant broke the law.

Taking the reasonable doubt standard to be in part a communication about the jurors' circumstances also matches some jury instructions in criminal proceedings. “A reasonable doubt,” according to New York courts, “is a doubt that *a reasonable person, acting in a matter of this importance*, would be likely to entertain because of the evidence that was presented or because of the lack of convincing evidence.”¹⁸²

Why should there be an outright belief requirement? If the standards required a credence below 1, then they would permit the jury to gamble on the outcome of a case. This is what the statistical evidence problem reveals. “Tolerating a system,” Tribe wrote, “in which perhaps one innocent man in a hundred is erroneously convicted despite each jury’s attempt to make *as few mistakes as possible* is in this respect vastly different from instructing a jury to *aim* at a 1% rate (or even a .1% rate) of mistaken convictions.”¹⁸³

¹⁸⁰ *Lampe v. Franklin Am. Tr. Co.*, 96 S.W.2d 710, 723 (Mo. 1936) (quoting *Rouchene v. Gamble Constr. Co.*, 89 S.W.2d 58, 63 (Mo. 1935)).

¹⁸¹ This is also true within legal scholarship. L. Jonathan Cohen, for example, used a very different concept of “belief” from the one used here. See generally L. Jonathan Cohen, *Should a Jury Say What It Believes or What It Accepts?*, 13 *CARDOZO L. REV.* 465 (1991). On his concept, belief comes in degrees, like credences, and is distinguishable from a concept called *acceptance*, which does not come in degrees and requires active rather than passive cognition. See *id.* at 465-66, 476. It seems that his notion of acceptance is in many ways like contemporary epistemologists’ notion of outright belief, and his notion of belief is more like what contemporary epistemologists would call a credence.

¹⁸² *Criminal Jury Instructions & Model Colloquies: Final Instructions*, NYCOURTS.GOV 9 (emphasis added), http://www.nycourts.gov/judges/cji/5-SampleCharges/CJI2d.Final_Instructions.pdf [https://perma.cc/6Z7D-YG9B].

¹⁸³ Tribe, *supra* note 7, at 1374 n.143. See also Nesson, *supra* Part I.B.2.

Allowing jurors to make bets permits the state to treat false convictions as something other than mistakes.¹⁸⁴ Suppose a juror with a 0.96 credence that a defendant broke the law were to convict him of a crime, and suppose that it was later revealed the defendant was innocent. One thought is that it is impossible to say the juror was wrong in her judgment. She might say, “I found there was a 96 out of 100 chance the defendant was guilty. Out of a hundred people convicted based on that assessment, four would be innocent. Nothing was wrong with my factual conclusion. This defendant must just be one of the four.” The legal system’s architects might say the same thing. Wrongful convictions would not necessitate any soul-searching; no questions would need be asked about what went wrong, for the simple reason that nothing did go wrong.

It is hard to see how this response could be just. As Andrea Roth has argued, “if the public understands that jurors are merely betting on guilt rather than being personally convinced before condemning a potential innocent, it will perceive the system as under-valuing individual dignity.”¹⁸⁵ And, crucially, the public’s judgment would be correct. A system that did not treat false convictions as mistakes would be expressly permitting—even endorsing—the violation of the individual rights of a few falsely convicted persons to secure other public interests. A paradigmatic individual dignity violation.

The intuition that this is wrong may be less strong in civil cases than criminal ones, but it may still be there.¹⁸⁶ Civil penalties,

¹⁸⁴ See Nesson, *supra* note 7, at 1361; Roth, *supra* note 30, at 1164 (citing WILLIAM WILLS, AN ESSAY ON THE PRINCIPLES OF CIRCUMSTANTIAL EVIDENCE 6 (Sir Alfred Wills ed., 5th ed. 1905)).

¹⁸⁵ Roth, *supra* note 30, at 1165.

¹⁸⁶ This is not to deny alternative theories of liability in the civil context, which relax traditional requirements. One example is market share liability, which courts have adopted in the medical malpractice tort context. Market share liability relaxes the traditional *causation* requirement in cases when certain facts are alleged. For example, in *Sindell v. Abbott Laboratories*, a patient alleged harm due to a synthetic drug. 607 P.2d 924, 925 (Cal. 1980). The drug was sold by several different defendant companies, all of whom, the plaintiff had shown, acted wrongfully in selling the drug. *Id.* at 936-38. But because the drug formula was fungible, the plaintiff could not trace the drugs to any one company. *Id.* at 936. The court permitted the plaintiff to recover against all of the companies who had engaged in the wrongful conduct, each according to its market share. *Id.* at 937-38. See also David A. Fischer, *Products Liability—An Analysis of Market Share Liability*, 34 VAND. L. REV. 1623, 1623-25 (1981). Another example is *Summers v. Tice*, 199 P.2d 1, 5 (Cal. 1948), a famous case in which the court placed the burden of proving

too, can cause suffering when wrongfully imposed. Indeed, a serious civil penalty, like confinement, might impact a person more severely than a minor criminal penalty, like a small fine, and the state is still the one imposing the mandate. If the state is to take these acts against people, the law is saying, then it should be sure that they broke the law.¹⁸⁷ Or the reason might not turn on the consequences at all. It might simply be that holding someone responsible for some act requires—as a matter of morality or longstanding social custom—a belief that she actually committed the act.¹⁸⁸

causation on two defendants, each of whom had acted wrongfully in precisely the same manner, and one of whose conduct had surely caused the injury. *See also* Judith Jarvis Thomson, *The Decline of Causation*, 76 GEO. L.J. 137, 137-39 (1987). In these cases, courts chose not to make causation a part of the claim that the plaintiff must prove. The preponderance standard in these cases applied only to the other traditional elements of the claim—for example, that the defendants acted wrongfully, and that the wrongful conduct of at least one of them caused an injury.

¹⁸⁷ The outright belief requirement thus also applies to the *clear and convincing evidence* standard, which is an intermediate civil standard: higher than the preponderance standard and lower than the reasonable doubt standard. *See generally* Emily Sherwin, *Clear and Convincing Evidence of Testamentary Intent: The Search for a Compromise Between Formality and Adjudicative Justice*, 34 CONN. L. REV. 453 (2002). Clear-and-convincing-evidence applies in certain civil situations of particular importance, such as the involuntary commitment of individuals for psychiatric treatment. *See Addington v. Texas*, 441 U.S. 418, 431 (1979) (“Having concluded that the preponderance standard falls short of meeting the demands of due process and that the reasonable-doubt standard is not required, we turn to a middle level of burden of proof [‘clear and convincing’ evidence] that strikes a fair balance between the rights of the individual and the legitimate concerns of the state.”).

¹⁸⁸ This may be the road taken by Lara Buchak. *See generally* Buchak, *supra* note 26. Buchak’s argument is specifically concerned with how reasonable notions of *blame* rely on an outright belief that a person committed the blameworthy act. Blame appears unjustified if we think there is merely a high chance the person committed the blameworthy act. *Id.* at 299-300. This explanation is highly credible for criminal cases, but it may be harder to sustain in the civil context, when courts generally do not speak of people being *culpable* or *blameworthy* for having feasted a tort or broken a contract. Indeed, it is commonly argued in legal scholarship that tort law, and perhaps contract law, should focus above all on the proper allocation of social costs and incentives—not the proper allocation of blame. *See, e.g.*, Guido Calabresi, *Some Thoughts on Risk Distribution and the Law of Torts*, 70 YALE L.J. 499, 499-500 (1961); Richard A. Posner, *The Law and Economics of Contract Interpretation*, 83 TEX. L. REV. 1581, 1591-92 (2005). But Buchak also suggests that “belief is ineliminable from our best theories about the norms associated with *holding each other responsible*.” Buchak, *supra* note 26, at 296 (emphasis added). This may be a promising alternative path. Even if courts do not blame people for breaching contracts in the sense of holding them criminally culpable, they still seem to hold them responsible. At the least, they factually assert that defendants broke

A similar question is how the requirement came to be part of the law. If it takes modern epistemology to reveal that an outright belief requirement is possible, then there is an apparent mystery as to how it became a part of standards of proof written long ago. This is also a question for further development, but, as a first pass, it might be that these assumptions about beliefs, while usually unarticulated, have long been common-sense ideas in the back of people's minds, governing people's social practices, including those of the people who helped develop the standards of proof. The standards' developers might have found plausible the idea that jurors should not be making bets on whether defendants have broken the law. But this, too, is an area for further thought.

C. Operationalizing the Outright Belief Requirement

Although at first it may appear to be a clean departure from earlier explanations of the standards of proof, an outright belief requirement is commensurable with most of their premises. For example, while the credence account is insufficient by itself to explain the standards of proof, jurors might still reason in a broadly Bayesian fashion consistent with an outright belief requirement. A juror might weigh each piece of evidence, determining how much it supports or undermines the plaintiff's case, remaining unsure about whether the defendant broke the law throughout most of the trial. But by the end of the trial, something else must happen—some mental alchemy must turn the juror's reasoning into a credence of 1. A broadly Bayesian methodology might be a substantial part of her reasoning; it just cannot be the only part. Similarly, jurors subject to an outright belief requirement might tend to reason in ways broadly consistent with the relative plausibility account. They may be persuaded that the only live possibilities as to what happened in the case are the stories each side is telling. They might thus engage in a process of inference to the best explanation. But in order to convict the defendant, the

the law. Cf. H. L. Ho, *Re-Imagining the Criminal Standard of Proof: Lessons from the 'Ethics of Belief'*, 13 INT'L J. EVID. & PROOF 198, 203 (2009) ("To give a guilty verdict is, first, to assert that the accused did commit the crime."). Likewise, to give a liable verdict is to assert that the defendant did breach the contract.

inference they end up with must be an outright belief that the defendant broke the law.

Yet the argument does have a practical consequence.¹⁸⁹ It means that some courts have been giving improper instructions to juries in civil cases. An instruction that tells jurors that “[a] party must persuade you, by the evidence presented in court, that what he or she is required to prove is *more likely to be true than not*,”¹⁹⁰ is telling jurors that all that is required to find the defendant liable is for them to believe it is 0.51 likely that the defendant broke the law. The inescapable conclusion, if one accepts this Article’s argument, is that these instructions are mistaken on the law.

Other common instructions on the preponderance standard do not make this error. These instructions tell jurors “to ‘think about an old-fashioned balanc[ing] scale’ with ‘all the believable evidence favorable to the plaintiff in one pan,’” and the evidence favoring the defendant in the other pan.¹⁹¹ “If the scales tip, even slightly, to the plaintiff’s side,” then the plaintiff wins.¹⁹² Some have criticized weight-based instructions on the grounds that they are vague and fail to clarify much about the preponderance standard.¹⁹³ But in this respect—a lexically prior one—they are superior. They do not incorrectly describe the law.

¹⁸⁹ The Article does not consider normative criticism of the standards of proof, but the view it sketches might provide a frame for doing so. For example, if the law really views the stakes of criminal cases as categorically more important than the stakes of civil cases—hence it sets the standard of proof in criminal cases higher than civil ones—one normative question is, is the law right? As briefly noted, it may be that some civil remedies are more consequential than some criminal penalties. If this is true, then that might counsel adapting the standards accordingly. Perhaps it would require that the reasonable standard be used in civil cases when penalties of a certain consequence were in play. Or, more radically, it might counsel overhauling the standards altogether. The point is not to endorse this or any other normative criticism. Rather, it is to show how the Article’s account of the standards of proof could prompt such kinds of criticisms.

¹⁹⁰ See Leubsdorf, *supra* note 4, at 1573 (emphasis added) (quoting JUDICIAL COUNCIL OF CAL., CIVIL JURY INSTRUCTIONS 200 (2005)). See also JUDICIAL COUNCIL OF CAL., CIVIL JURY INSTRUCTIONS 40 (2020), available at https://www.courts.ca.gov/partners/documents/Judicial_Council_of_California_Civil_Ju_Ju_Instructions.pdf [<https://perma.cc/28WX-QHZ9>].

¹⁹¹ *Leubsdorf, supra* note 4, at 1572 (citation omitted).

¹⁹² *Id.*

¹⁹³ See, e.g., *id.* (“This metaphor has the aesthetic advantage of being just as unclear as, and even quaintier than, the term it explains.”).

The balancing metaphor states that the jury should simply weigh the evidence to determine what happened.¹⁹⁴ To see how this would work, one could apply the metaphor of the balancing scale to the hiker and her friend listening to what might be a woodpecker tapping a nearby tree. Suppose the hiker's friend tells her that the sound is probably a man-made noise. The hiker might weigh this evidence, putting it in the pan on one end of the scale. Then the hiker might look at the tree and see that a woodpecker has previously bored holes on this side of the tree. The hiker would take this evidence and put it in the pan on the other end of the scale. After looking around some more, the hiker might see a woodpecker's nest higher up on the tree and add that evidence to the pan on the scale. After reasoning in this matter, it seems, a reasonable person could clearly wind up with an outright belief that it is a woodpecker making the tapping noise.

Moreover, the metaphor is apt. A balancing scale gets to the heart of what the preponderance standard is asking jurors to do in giving equal right to both sides. It is different from criminal cases, in which the prosecutor bears a far greater burden than the defendant. The metaphor is correctly describing the law's view of the circumstances in which they are making up their minds. While they might be vague, metaphoric instructions on the preponderance standard are preferable to incorrect ones.

The difference between the preponderance and reasonable doubt standards is that in preponderance cases jurors should allow themselves to be more willing to form an outright belief. They should be less doubting. They should treat fewer alternatives as "live."¹⁹⁵

¹⁹⁴ One could argue that the metaphor is asking jurors to determine whether it is more probable than not that the defendant did what has been claimed. Under this interpretation, it would be no different from instructions that use expressly probabilistic language in instructing the jury. But while balancing scales might be *incremental*, they are not probabilistic, which may be what makes the difference.

¹⁹⁵ See Greco, *supra* note 24, at 186. The matter of jury instructions, too, is an occasion for further research. Historically, courts may have been more willing to use concepts like belief when instructing jurors on the preponderance standard. It has been argued that in the latter part of the nineteenth century, courts found that the preponderance standard required evidence to "convince or satisfy jurors . . ." See Leubsdorf, *supra* note 4, at 1614. Variations on such older instructions may offer alternative possibilities for courts to consider, which could aptly describe the preponderance standard without falling back on probabilistic explanations.

It also seems to be an intuitively plausible account of what it is *like* to be a juror in a civil case. From memory, that was the experience of this Article's author when once serving as a mock juror in law school, before I thought about these questions. The task, it seemed, was not to determine the probability that the defendant did it, or to see whether the plaintiff's specific story was better than the defendant's specific story. Rather, it was to answer a simpler question: did the defendant do what the plaintiff claimed?

CONCLUSION

An outright belief requirement plausibly resolves an enduring mystery in the standards of trial proof. It explains why statistical evidence is insufficient to license liability, because it cannot cause a reasonable person to believe outright that the defendant broke the law. And it resolves the conjunction problem, because 1 is the only credence for which proof of each non-dependent element is the same as proof of their conjunction. Furthermore, it systematizes one of the great discoveries about trials that the statistical evidence and conjunction problems have unearthed. Lawyers and laypersons alike are uncomfortable with juries making bets based on trial evidence.

The standards of proof are, in part, communications to the jury about the circumstances in which they are making up their minds. The circumstances of liability prescribed by the preponderance standard are, in the law's view, *less significant* for defendants than the circumstances of guilt prescribed by the reasonable doubt standard. Thus, jurors must be more reluctant to form an outright belief that the defendant broke the law in criminal cases.

