PROBABILITY, CONFIDENCE, AND THE “REASONABLE JURY” STANDARD

Luke Meier*

INTRODUCTION

I. FOUNDATIONAL CONSIDERATIONS
   A. “Direct” and “Circumstantial” Evidence
   B. Litigation “Burdens”
      1. The Burden of Persuasion
      2. The Burden of Production
   II. PROBABILITY AND THE BURDEN OF PRODUCTION
   III. CONFIDENCE AND THE BURDEN OF PRODUCTION

A. The Blue Bus and Gatecrasher Hypotheticals:
   Descriptive Questions
B. The Blue Bus and Gatecrasher Hypotheticals:
   Normative Explanations
C. Professor Neil Cohen’s Confidence Principle
D. Confidence: Burden of Production Only or
   Burden of Persuasion Also?

CONCLUSION

INTRODUCTION

Rule 56 of the Federal Rules of Civil Procedure provides that a federal court “shall grant summary judgment if the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law.”1 In 1986,

---

* Associate Professor of Law, Baylor Law School. J.D., University of Texas School of Law. B.S., Kansas State University. This paper is part of a larger project that has benefited from the feedback of numerous individuals deserving of thanks: Kevin Clermont, Ken Klein, Jill Lens, Clyde Martin, Charles Nesson, Jerry Powell, Rory Ryan, Larry Solum, Jeff Stempel, Peter Tillers, Maggie Wittlin, Jim Wren, and Richard Wright. I am also grateful for the research help provided by the following individuals: Ryan Jamison, Erin Hamilton, Emily Oglesby, Andrea Palmer, and Kim Pearson. All errors are mine.

1 FED. R. CIV. P. 56(a).
the Supreme Court decided the infamous trilogy of cases that ultimately elevated summary judgment to greater prominence within the pantheon of litigation procedure. In the aftermath of this trilogy, summary judgment has received considerable academic attention from legal commentators. Numerous empirical studies have documented the decreasing frequency of civil trials in federal court, and commentators often attribute this trend to the modern, vigorous use of summary judgment. The constitutionality of summary judgment under the Seventh Amendment’s guarantee of a right to trial has been the subject of considerable debate. More recently, the relationship between

---


3 See Kevin M. Clermont, Litigation Realities Redux, 84 NOTRE DAME L. REV. 1919, 1941-42 (2009) (discussing academic attention to summary judgment).


summary judgment and the “plausibility” standard for analyzing motions to dismiss has been considered. Extensive academic discussion of these various aspects of summary judgment has triggered a general discussion as to whether summary judgment is a valuable component of federal civil procedure.

Despite the substantial interest in summary judgment, it is still misunderstood. Much of the blame for this misunderstanding can be attributed to the standard articulated by the Supreme Court for determining when summary judgment is appropriate.

In *Anderson v. Liberty Lobby, Inc.*, the Court held that summary judgment under Rule 56 is to be measured by the “reasonable jury” standard. To understand the difficulties created by the reasonable jury gloss on Rule 56 (the term “reasonable jury” does not appear within the text of Rule 56), it is necessary to distinguish between the concepts of “probability” and “confidence” as they apply to disputed questions of fact in a litigation context. This Article explores both concepts. In brief, “probability” refers to...
the likelihood that a litigated fact is true, and “confidence” refers to the margin of error associated with a particular probability assessment. Stated differently, the concept of probability requires an estimate as to the likelihood of a given fact being true; the confidence inquiry asks how sure one is in the accuracy of that probability estimate.

While the probability concept is easily understood, the confidence principle is more elusive, at least for those without a mathematical background. The best explanation and discussion of the confidence principle, as it applies in the litigation context, is an academic debate involving the famous “blue bus” and “gatecrasher” hypotheticals posed by Professors Laurence Tribe and Jonathan Cohen, respectively. Unfortunately, insights from these extended academic discussions have been largely confined to academia. As such, the confidence principle remains mostly unrecognized as a distinct concept within conventional procedural doctrine, even though judges sometimes engage in this type of analysis during the course of litigation.

The reasonable jury standard is a substantial impediment to the widespread identification of the confidence principle within working legal doctrine. The reasonable jury standard implies that the judge’s analysis at summary judgment essentially replicates the jury’s analysis at trial. In reality, however, the judge’s analysis at the summary judgment stage is materially different than the fact-finder’s analysis at trial. A jury is concerned primarily, if not exclusively, with the probability question. Conversely, in deciding whether to allow a case to survive summary judgment and proceed to trial, a judge analyzes the record evidence by considering both probability and confidence.

The reasonable jury standard emphasizes the relationship between the judge and jury with regard to the *probability* question and adequately captures the notion that the jury has the primary responsibility in deciding the probability question. Ultimately, the reasonable jury standard obscures the fact that judges often grant summary judgment (usually against a plaintiff) because the dearth of evidence on a material fact means that insufficient confidence can be had in *any* probability assessment as to that

---

10 See infra notes 96-98, 130-32 and accompanying text.
material fact. Further, when the confidence principle goes unidentified as the true basis for a summary judgment, it suggests that the judge has assumed a greater role in assessing the probability question than the “reasonable jury” standard permits. In reality, the judge has simply used an analysis that is not adequately captured by the reasonable jury phrase.

The failure to distinguish between probability and confidence has resulted in a litany of unfortunate misunderstandings regarding federal pre-trial procedure. Distinguishing between probability and confidence is the key to understanding the analysis required by the plausibility standard introduced in the controversial case of *Bell Atlantic Corp. v. Twombly*. In addition, once the difference between probability and confidence is appreciated, it diminishes the argument that modern pre-trial procedure violates the Seventh Amendment. Also, and perhaps most importantly, recognizing the difference between probability and confidence permits a new understanding of the Supreme Court’s pivotal decision in *Matsushita Electric Industrial Co. v. Zenith Radio Corp.* Pursuant to this new interpretation of *Matsushita*, under which the defendants’ summary judgment was based on a confidence analysis rather than a probability analysis, *Matsushita* is no longer the game-changing case signaling a new era in summary judgment law. Rather, the case’s importance is mostly limited to the unique context of antitrust law cases. The summary judgment “revolution” of 1986, it turns out, was the result of a misunderstanding deriving from the failure to appreciate the difference between probability and confidence.

The explanation and discussion of the various consequences deriving from the inability to clearly distinguish between probability and confidence is an immense topic and will be developed fully in other articles. The objectives for this Article are more limited, and two-fold. One, this Article will demarcate the difference between probability and confidence, as those concepts relate to disputed questions of fact in a litigation context. In pursuing this goal, the Article will rely heavily on previous academic discussions regarding this topic. Other academics have

---

12 475 U.S. 574 (1986).
13 See id. at 585-88.
also identified the distinction between probability and confidence as they pertain to the litigation context. This previous work will be used and, in some respects, expanded upon to address shortcomings in the previous discussion. Two, this Article will attempt to explain why the difference between probability and confidence has not yet been incorporated into mainstream procedure doctrine. Over twenty-five years ago, Professor Neil Cohen explained that delineating between probability and confidence was the key to understanding the “paradox” of the blue bus and gatecrasher hypotheticals posed by Professors Laurence Tribe and Jonathan Cohen.¹⁴ Over two decades later, however, it appears that Professor Cohen’s insights were relevant only within the confines of the ivory tower. In reality, distinguishing between confidence and probability sheds light on a variety of contemporary—and problematic—federal procedural issues. The problem, however, is that the “reasonable jury” interpretation of Rule 56 makes it extremely difficult to distinguish between confidence and probability. The second objective of this Article is to demonstrate that the reasonable jury standard has this effect.

The organization of the remainder of the Article is as follows: Part I discusses the important background topics of direct and circumstantial evidence, as well as the burdens of persuasion and production. Part II addresses the probability analysis and how this analysis relates to the relationship between judge and jury. Part III develops the confidence principle, demonstrating that (1) the confidence principle resolves the difficult questions raised by the infamous blue bus and gatecrasher hypotheticals, and (2) the confidence principle is necessarily part of a judge’s analysis at the summary judgment stage. The Article concludes with a discussion of how the reasonable jury gloss on Rule 56 makes the identification of the confidence principle more difficult.

I. FOUNDATIONAL CONSIDERATIONS

As an introduction to the topics discussed in this Article, two foundational topics will be briefly discussed in this section. The first is the difference between direct and circumstantial evidence.

The second is the difference between the burden of production and the burden of persuasion. Clearly distinguishing between these two “burdens” is imperative to understanding the ideas developed in this paper.

A. “Direct” and “Circumstantial” Evidence

Evidentiary scholars have long recognized the distinction between direct and circumstantial evidence.\(^{15}\) Simply put, direct evidence is evidence that, if credible to the jury, proves the material fact in the litigation.\(^{16}\) For example, if there is a dispute involving the material fact of whether a defendant stopped at a stop sign before an automobile accident, trial testimony from a witness that he saw the defendant stop at the stop sign would be direct evidence of this material fact. Circumstantial evidence, however, is evidence that, if believed by the jury, does not “directly prove” the material fact to the litigation, but rather supplies an inference that the material fact occurred.\(^{17}\) In the above example, testimony from a trial witness that she had ridden with the defendant hundreds of times and that he always stopped at the

\(^{15}\) See Chris Reed & Glenn Rowe, A Pluralist Approach to Argument Diagramming, 6 LAW, PROBABILITY & RISK 59, 61-64 (2007) (explaining renowned evidentiary scholar John Wigmore’s categorization of the types of evidence, which was based in part upon the difference between direct and circumstantial evidence); ALEXANDER M. BURRILL, A TREATISE ON THE NATURE, PRINCIPLES AND RULES OF CIRCUMSTANTIAL EVIDENCE, ESPECIALLY THAT OF THE PRESUMPTIVE KIND, IN CRIMINAL CASES 4 (1856) (distinguishing between direct and circumstantial evidence). To say that the distinction between direct and circumstantial evidence has been drawn is not, of course, equivalent to saying that these two types of evidence are always treated differently by courts. On this separate point, context matters. This Article will argue that the difference between direct and circumstantial evidence can matter because a party with direct evidence on a material fact is unlikely to suffer a summary judgment on that point, while the same cannot necessarily be said for a party relying only upon circumstantial evidence. This does not necessarily mean, of course, that a party relying only upon circumstantial evidence will lose. See, e.g., Desert Palace, Inc. v. Costa, 539 U.S. 90, 99 (2003) (explaining that plaintiffs in civil cases and the prosecution in criminal cases are permitted to prove their case by direct or circumstantial evidence).

\(^{16}\) See DENNIS D. PRATER ET AL., EVIDENCE: THE OBJECTION METHOD 116 (2d ed. 2002) (explaining that with direct evidence “[n]o inferences are required between the proof itself and the ultimate fact to be established”); cf. WILLIAM WILLS, AN ESSAY ON THE RATIONALE OF CIRCUMSTANTIAL EVIDENCE 24-25 (1838) (reasoning that direct evidence requires an inference only “as the thunder follows the flash”).

\(^{17}\) See PHIFSON ON EVIDENCE § 4 (John Huxley Buzzard et al. eds., 12th ed. 1976) (explaining the additional inference required by circumstantial evidence).
stop sign in question is circumstantial evidence that the defendant stopped at the stop sign on the date of the accident. Even if the jury believed the testimony of the witness (and thus believed that the witness had, in fact, ridden before with the defendant and that the defendant always stopped at the stop sign), an additional inference is necessary to prove the material fact—because the defendant stopped at that particular stop sign hundreds of times prior to the accident, the defendant stopped at the stop sign on the date of the accident.

Algebraic symbols can depict the difference between direct and circumstantial evidence. The material fact is represented as Fact M. Direct evidence of Fact M is testimony from a witness that he observed Fact M. This testimony is represented by Testimony M. The question for a jury with regard to direct evidence is whether Fact M can be deduced from Testimony M. This relationship can be depicted algebraically: Testimony M → Fact M. In order to deduce Fact M from Testimony M (represented by the arrow in the algebraic equation), the jury must find Testimony M credible. This means that the jury must conclude: (1) the witness perceived the event accurately, (2) the witness remembers the event accurately, (3) the witness is honestly testifying, and (4) the witness is able to accurately communicate his or her honest recollection of this event.18

Depicting circumstantial evidence through algebraic symbols makes clear the additional inference required for this type of evidence. Again, the material fact in the litigation is represented as Fact M. With circumstantial evidence, however, the trial testimony is not about a direct observation of Fact M by the witness, but rather about a fact that provides the basis of an inference for Fact M. This separate fact is represented as Fact X, with the testimony (or direct evidence) of Fact X represented as Testimony X. Thus, the probative force of circumstantial evidence is based on two inferences:

Testimony X → Fact X → Fact M. First, the jury must find credible the witness’s testimony that “X” occurred. This means that the jury must believe that the witness perceived the event accurately, remembers correctly, is honest, and is accurately conveying this information. But, an additional inference is required in order for Testimony X to be probative of Fact M: the jury must find some relationship between the existence of Fact X and Fact M.19

The work of a jury in deciding whether Fact X can be inferred from Testimony X (or Fact M from Testimony M) is similar to, but slightly different from, the work of a jury in deciding whether Fact M can be inferred from Fact X.20 Obviously, both inferences require a jury to consider after-the-fact evidence in determining whether Fact M actually occurred in the real world. None of the jurors will have directly observed Fact M because, under modern law, any juror claiming to have witnessed Fact M is precluded from jury duty.21 In this sense, direct and circumstantial evidence are similar in that they both require a jury to evaluate what really happened at a particular location and at a particular time based solely on information provided in a different location (namely, a courtroom) at a later point in time (the trial). From this perspective, the line between circumstantial and direct evidence is

---

19 Circumstantial evidence can involve multiple layers of inferences. For instance, if the material fact is Fact M, and a party wishes to prove Fact M by Fact X, the party might wish to prove Fact X with testimony about Fact W, under the belief that the existence of Fact W implies the existence of Fact X. This evidence is charted algebraically as Testimony W → Fact W → Fact X → Fact M. For a wonderful discussion of a hypothetical involving circumstantial evidence whose probative value is based on multiple inferences, see Richard D. Friedman, A Diagrammatic Approach to Evidence, 66 B.U. L. Rev. 571, 591-611 (1986). By carefully depicting the ubiquity of probability analysis based on “inferences from inferences,” Friedman does serious damage to the oft-quoted proposition that evidence based on “an inference upon an inference” is not even relevant and thus not admissible. See id. at 607 n.56 (discussing the admissibility of evidence requiring multiple inferences).


21 See generally John Marshall Mitnick, From Neighbor-Witness to Judge of Proofs: The Transformation of the English Civil Juror, 32 Am. J. Legal Hist. 201, 207 (1988) (stating that a juror with personal information regarding the facts of a case will be excluded from duty).
blurred because both belong to the larger category of after-the-fact evidence of a prior event.\(^2\)

\(^{2}\) See 1A John Henry Wigmore, Evidence in Trials at Common Law § 24, at 944 n.5 (Peter Tillers rev. 1983) (examining a variety of perspectives on the difference between direct and circumstantial evidence). On this point, consider the enlightening debate between Professors Charles Nesson and Neil Cohen over Nesson’s “card hypothetical.” In this hypothetical, Professor Nesson first takes a playing card from a deck and then quickly flashes the front of the card to a witness. See Charles Nesson, The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts, 98 Harv. L. Rev. 1357, 1361 (1985); cf. The Colbert Report (Comedy Central television broadcast Jan. 24, 2008), available at http://thecolbertreport.cc.com/videos/bpncyw/charles-nesson (staging an intervention to Professor Nesson’s “gambling problem”). The question to be answered is whether the pulled card was a king. The witness believes that the card that he quickly glimpsed was not a face card and thus was not a king. We could say that the witness has “direct evidence” on the material fact of whether the pulled card was a king. Next, Nesson takes a new card from the deck and, without showing it to the “witness,” asks whether the card is a king. The witness does not have any direct observation of the type of card that was pulled. So, in answering the question of whether the card is a king, the “witness” (actually, “expert” would probably be a more accurate description) relying on statistical probabilities would obviously be forced to “guess” that the blind card is not a king. In this instance, we could say that the “witness” has circumstantial evidence, because the witness is basing his estimate that the card is not a king on an inference from another fact (here, the number of kings in a normal deck) and not on direct observation. Nesson uses this example as the basis of his theory that direct observation, such as with the first witness, should be favored over statistical circumstantial evidence, because the former represents information about an actual event while the latter represents information about the evidence submitted at trial. According to Nesson, the legitimacy of the civil trial process is undermined when trials are perceived as being resolved on the basis of the evidence submitted at trial rather than the actual events that occurred in the real world. Professor Neil Cohen persuasively critiques Professor Nesson’s hypothetical. See generally Neil B. Cohen, The Costs of Acceptability: Blue Buses, Agent Orange, and Aversion to Statistical Evidence, 66 B.U. L. Rev. 563 (1986). Professor Cohen states: “In a trial, after all, the jury itself does not catch a glimpse of the card in question; rather a witness testifies that he or she caught a glimpse of the card.” Id. at 566; see also Richard K. Greengstein, Determining Facts: The Myth of Direct Evidence, 45 Hous. L. Rev. 1801, 1820 (2009) (“That is, all evidence tells stories, and the construction of those stories, along with concomitant judgments about their relationship to the truth, always depends on inferential reasoning—for ‘direct’ evidence every bit as much as for ‘circumstantial’ evidence.”). Professor Cohen has correctly recognized that, at a certain level, the difference between “direct” and “circumstantial” evidence is artificial, because in either instance the jury has to surmise what happened at an earlier date, in a different location, based on what is said or presented to them in the courtroom. See Cohen, supra, at 566. Professor Nesson, however, has also recognized an intuitive difference between a jury’s reasoning process required for “direct” versus “circumstantial” evidence, even though neither type is “direct” in the sense that the jury is able to actually observe the event in question. See Nesson, supra, at 1361-62. Professor Cohen acknowledges that Nesson’s thesis, including his card hypothetical, “has undeniable intuitive appeal.” See Cohen, supra, at 567.
Nevertheless, it is helpful to distinguish between direct and circumstantial evidence because the task of deciding whether Fact X can be inferred from Testimony X is different than the task of deciding whether Fact M can be inferred from Fact X. For direct evidence, the jury’s duty of assessing the probative value of the evidence primarily requires judgments as to what has transpired within the courtroom. Evaluating a witness’s credibility is a task that requires the fact-finder to rely primarily on direct observation of a witness in the courtroom.23 Alternatively, for circumstantial evidence, the task of evaluating whether the material fact can be deduced from the testified fact will more explicitly require the jury to rely on personal experiences from outside the courtroom.24 For instance, with regard to the circumstantial evidence from the witness who testified to riding with the defendant and that he always stopped at the stop sign, the jury has to consider whether drivers who have a history of driving safely continue to exhibit that characteristic going forward. The jury’s deductive analysis for circumstantial evidence requires far more reliance on personal, out-of-court experiences than what is necessary for weighing the credibility of a witness’s direct evidence.

23 See H. Richard Uviller, Credence, Character, and the Rules of Evidence: Seeing Through the Liar’s Tale, 42 DUKE L.J. 776, 779 (1993) (explaining that a witness’s credibility is “principally” established through “cross-examination and contradiction” during the trial process). But see Jeremy A. Blumenthal, A Wipe of the Hands, a Lick of the Lips: The Validity of Demeanor Evidence in Assessing Witness Credibility, 72 NEB. L. REV. 1157, 1159-60 (1993) (acknowledging the historical view that a witness’s demeanor at trial informs the jury about the credibility of that witness but challenging that notion based on empirical tests).

24 See Commonwealth v. Webster, 59 Mass. (5 Cush.) 295, 312 (Mass. 1850) (“Circumstantial evidence . . . is founded on experience and observed facts and coincidences, establishing a connection between the known and proved facts and the fact sought to be proved.”) (emphasis added). Along these lines, consider Judge Posner’s suggested reform of trial procedures in antitrust cases, as spelled out in Chapter X of his seminal book Antitrust Law. Posner’s suggestion involves the creation of a “chronological narrative” of the material facts to the dispute. See RICHARD A. POSNER, ANTITRUST LAW 283 (2d ed. 2001). This chronological narrative would then be submitted to the jury in “a writing that would constitute the basic trial record.” Id. To the extent that the lawyers to the case disagreed upon “basic facts,” live testimony in front of the jury would be required. Id. However, if the parties disagreed only upon the “inferences to be drawn from” agreed-upon basic facts, no live testimony would be necessary, and the jury would be expected to rely on their experience and common sense in determining whether the inferences could be drawn from these agreed-upon “basic facts.” Id. at 284.
Admittedly, the above explanation of the jury’s role in determining the probative force of Testimony X to Fact X, as compared to the jury’s role in determining the probative force of Fact X to Fact M, is inexact. In both instances, the jury will have to rely on experiences that occur both in the courtroom during trial and from past experience outside the courtroom. For example, how does a jury value the credibility of a witness who appears nervous? The juror must apply his or her past experiences with nervous people and lying people.25 Further, for circumstantial evidence, a juror might be assisted by the in-trial introduction of statistical information that demonstrates the relationship between Fact X and Fact M.26 That said, the task of determining the probative force of Testimony X to Fact X is different than the task of determining the probative force of Fact X to Fact M. Courts have recognized this distinction by conferring a privileged status to direct evidence. As explained below, a party with direct evidence on a material fact is almost always shielded from summary judgment.

B. Litigation “Burdens”

The terms “burden of production,” “burden of persuasion,” and “burden of proof” have historically been the source of frustrating misunderstandings.27 Some of the ambiguity involving these terms arises from the fact that “burden of proof” has sometimes been used to refer to the two more specific concepts now properly distinguished by the separate terms “burden of

25 See generally Blumenthal, supra note 23, at 1189-1200 (explaining that a juror’s ability to determine the credibility of a trial witness will depend in part on the juror’s preconceived notions regarding how honest people act and exploring, empirically, the problems with this model).


27 CHARLES T. MCCORMICK, HANDBOOK OF THE LAW OF EVIDENCE § 308, at 639 (1954) (describing the term “burden of proof” as one of the “slipperiest member[s] of the family of legal terms”).
persuasion” and “burden of production.” At times, “burden of proof” has been used in reference to the concept now associated with the term “burden of persuasion,” and, in other instances, “burden of proof” has been used in reference to the concept now associated with the term “burden of production.”

Even further, “burden of proof” has been used to refer to both concepts. For the most part, however, it seems that the legal community has mostly graduated from the “lamentable ambiguity of phrase and confusion of terminology under which our law has so long suffered,” and that the somewhat ambiguous term “burden of proof,” when it is still used, refers generically to both concepts now associated with the terms “burden of production” and “burden of persuasion.”

Even apart from this lingering ambiguity associated with the term “burden of proof,” there is still some confusion generated by the terms “burden of production” and “burden of persuasion.” The problem is that these phrases can be employed to address two entirely different distinctions. The first distinction addresses the trial court’s division of power between the judge and the jury. This is the manner in which the terms will be used in this Article. Occasionally, however, the terms “burden of persuasion” and “burden of production” are used to refer to the type of analysis that is expected by the decision-maker. In this Article, I will use different terms—“probability” and “confidence”—to make this distinction.

29 Id.
30 Id.
32 See, e.g., 21B WRIGHT & GRAHAM, supra note 28, § 5122 (using the term “burdens of proof” as a generic term encompassing the “burden of production” and the “burden of persuasion”) (emphasis added). But see Louis Kaplow, Burden of Proof, 121 YALE L.J. 738, 773 (2012) (using the term “burden of proof” rather than “burden of persuasion”).
33 Kaplow, supra note 32, at 773 (using the term “burden of proof” to refer to the requirement that a plaintiff prove each material fact to a certain probability, but not distinguishing between the respective role of the trial court judge and the jury in resolving this issue).
1. The Burden of Persuasion

As used in this Article, the term “burden of persuasion” refers to the showing a plaintiff must make in order to convince the jury to find for the plaintiff. The concept can be expressed from the perspective of the jury. The “burden of persuasion” is the standard a jury applies in determining whether it can find for the plaintiff. Accordingly, the term “burden of persuasion” acknowledges the division of power between the jury and the trial court judge. The “burden of persuasion” simply refers to the jury’s respective authority within the trial court proceedings.

This understanding of the burden of persuasion identifies nothing concerning the nature of the jury’s analysis in determining whether to find for the plaintiff. On this separate question, however, it is clear that a probability analysis is, at least, part of the burden of persuasion. A plaintiff cannot be expected to prove with absolute certainty that a past event, on which the defendant’s liability is premised, actually occurred.

Determinations of past events, however, cannot recreate those events with perfect knowledge. “Time is irreversible, events unique, and any reconstruction of the past is 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.”

---

34 In some instances, such as with an affirmative defense, the burden of persuasion might be on the defendant. See Schaffer ex rel. Schaffer v. Weast, 546 U.S. 49, 57 (2005) (“[T]he burden of persuasion as to certain elements of a plaintiff’s claim may be shifted to defendants, when such elements can fairly be characterized as affirmative defenses or exemptions.”). For purposes of simplicity and clarity, however, this Article will uniformly refer to the burden of persuasion as a burden of the plaintiff. See id. (stating that the “ordinary default rule” is that the plaintiff bears the burden of persuasion).

35 Actually, the burden of persuasion refers to the duty to convince the fact-finder, but this Article will speak in terms of “the jury” rather than “the factfinder” for ease of terminology.

36 See Cohen, supra note 14, at 390-91 (“[T]here is general agreement on two fundamental points. First, probabilistic techniques may be used to determine the likelihood of the facts supporting a defendant’s guilt or liability. Second, the plaintiff’s or prosecutor’s burden is satisfied when that probability exceeds a threshold value.”) (footnote omitted).

Thus, our system accepts that a plaintiff, in attempting to meet her burden of persuasion, can succeed despite the jury believing that there is a possibility that the facts needed for the plaintiff’s recovery did not, in fact, occur in the real world.

Different “tests” or “standards” reflect the degree of probability a jury must assign to each material fact needed for the plaintiff’s recovery. In most civil disputes, probability is measured under the “preponderance of the evidence” standard. Commentators generally agree that a preponderance of the evidence standard requires a plaintiff to convince the jury that the probability of each material fact is greater than 50%. This concept is also reflected in typical federal civil jury instructions.

38 See 21B WRIGHT & GRAHAM, supra note 28, § 5122, at 409 (“The normal burden of persuasion in a civil case requires only that party prove the fact by a ‘preponderance of the evidence.’”).

39 “[A] fact is material if it tends to resolve any of the issues that have been properly raised by the parties.” 10A CHARLES ALAN WRIGHT, ARTHUR R. MILLER & MARY KAY KANE, FEDERAL PRACTICE & PROCEDURE § 2725, at 419 (1998). See infra note 132 and accompanying text for a discussion of the conceptual difficulties that arise from the assumption that a plaintiff must prove each material fact by a preponderance of the evidence.

40 See Vern R. Walker, Preponderance, Probability and Warranted Factfinding, 62 BROOK. L. REV. 1075, 1076 n.5 (1996) (listing commentators (and judicial opinions) that associate the preponderance of the evidence standard with 50% probability). Some commentators have argued that the traditional probability model of the burden of persuasion is incomplete and have provocatively suggested an “inference to the best explanation” approach as a superior model. See Michael S. Pardo & Ronald J. Allen, Juridical Proof and the Best Explanation, 27 LAW & PHILOS. 223, 231 (2008); see also Kevin M. Clermont, Death of Paradox: The Killer Logic Beneath the Standards of Proof, 88 NOTRE DAME L. REV. 1061 (2013) (advocating a “fuzzy logic” approach to understanding the burden of persuasion). Essentially, under this approach (and at the risk of simplification), jury’s do not (and should not) determine the probability of the plaintiff’s version of the disputed facts but instead compare the plaintiff’s explanation of the relevant evidence to the defendant’s explanation of the relevant evidence. See Michael S. Pardo, Second-Order Proof Rules, 61 FLA. L. REV. 1083, 1093 (2009) [hereinafter Second-Order] (concluding that “some comparison with the defendant’s case is necessary” under the preponderance of the evidence standard). This Article will employ the traditional probability conception of the burden of persuasion.
defining a “preponderance of the evidence.” Because the preponderance of the evidence standard requires a finding of greater than 50% probability, it is sometimes referred to as the “more likely than not” standard. The preponderance of the evidence standard for the burden of persuasion thus serves an additional objective, which is to instruct the jury how to proceed if they believe that the probability of a material fact necessary for the plaintiff’s case is exactly 50%. In this instance, the jury is to render a verdict against the plaintiff.

The preponderance of the evidence standard, however, is not the only measure of the probability question within the burden of persuasion. In some cases, the plaintiff must convince the jury to assign a higher probability to each—or at least some—of the material facts. Thus, for certain types of civil cases, the plaintiff’s burden of persuasion is “clear and convincing evidence.” Outside the context of civil cases, the familiar “beyond a reasonable doubt” burden of persuasion applies to criminal cases. Although the preponderance of the evidence standard is understood to mean more than 50%, no similar numerical value is assigned to either the “clear and convincing evidence” or “beyond a reasonable doubt” standards.

---

41 See, e.g., Federal Civil Jury Instructions of the Seventh Circuit § 1.27, at 34 (2009) (defining a preponderance of the evidence standard as requiring the jury to believe that the fact is “more probably true than not true”).

42 See Second-Order, supra note 40, at 1091-93 (discussing the preponderance standard with reference to the “more likely than not” approach). But see McCauliff, supra note 37, at 1330 (describing the results of a survey of judges which indicate that, although most judges consider “preponderance of the evidence” to be equivalent to “more likely than not,” a few judges consider these standards to be different).


44 See Kaplow, supra note 32, at 758 (noting that “ties” regarding probability go to a defendant). But see id. n.34 (suggesting that actual ties are unlikely (or even impossible) but that a jury might apply the tie-breaking function to any probability assessment close to 50%).


46 See In re Winship, 397 U.S. 358 (1970) (interpreting the Due Process Clause to require the beyond a reasonable doubt standard in criminal cases).

47 See Emily Sherwin, Clear and Convincing Evidence of Testamentary Intent: The Search for a Compromise Between Formality and Adjudicative Justice, 34 Conn. L. Rev. 453, 462 (2002) (stating the level of probability required by the clear and
requires the fact-finder to assign a higher probability than the preponderance of the evidence standard, and the clear and convincing standard is understood to fall somewhere between the reasonable doubt and preponderance of the evidence standards. As no exact percentages are assigned, the jurors are presumably left to their own devices in associating the exact probability required of these standards.

Some uncertainty exists as to whether the probability analysis is the only measure by which a jury determines whether a plaintiff has met the burden of persuasion. Some commentators argue and some empirical evidence suggests that the burden of persuasion requires not only that a plaintiff convince the jury as to the probability of the material fact in the litigation, but also that the jury has a certain confidence in its probability conclusion. The confidence concept will be explored in depth in Part III of this Article; at that time, the question as to whether the burden of persuasion includes a confidence analysis will be addressed.


Because no exact numerical probability is assigned to the “clear and convincing evidence” and “beyond a reasonable doubt” standards, the frequent comment that the “burden of persuasion” serves a “tie-breaking” function does not necessarily translate outside the context of the preponderance of the evidence standard. See Richard D. Friedman, Dealing with Evidentiary Deficiency, 18 CARDOZO L. REV. 1961, 1970 (1997) (equating the burden of persuasion with the tie-breaking function). Because no exact mathematical number is assigned to these measures of probability, the jury is presumably left to their own devices in determining the exact probability associated with these numbers, and the odds of a juror concluding that the probability of a material fact necessary for the plaintiff's (or prosecution's) case falling directly on the number they associated with the respective standard seems much more remote than when dealing with the 50% number. Cf. Kaplow, supra note 32, at 758 n.34 (suggesting that ties under the preponderance of the evidence standard are remote). Even if this were to occur, however, a verdict for one party or the other would not really be a “tie-breaker” because, according to the jury, the plaintiff's interpretation of the facts is more likely than the defendant's interpretation, but simply not to the level required under that particular standard. There is no “tie” in such a case.

See generally infra Part III.C.
2. The Burden of Production

As used in this Article, the term “burden of production” refers to the evidentiary record a party must compile to be entitled to a jury determination on contested material facts.\(^{51}\) The burden of production can be contrasted with the burden of persuasion by noting the different entity to which each burden is directed. As explained above, the burden of persuasion is the duty to convince a jury. The burden of production, on the other hand, is the duty to convince a judge to allow a case to reach the jury.\(^{52}\) If a party has not met her burden of production, the judge will enter a judgment against her.\(^{53}\)

Rules 50 and 56 of the Federal Rules of Civil Procedure provide the procedural bases for an entry of judgment resulting from failure to meet the burden of production. Rule 56 allows a trial court judge to enter summary judgment against a party if discovery reveals that “there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law.”\(^{54}\) Rule 50 allows for a judgment “as a matter of law” against a party after that party “has been fully heard on an issue during a jury trial and the court finds that a reasonable jury

\(^{51}\) In this Article, the term “burden of production” is used as a defensive, rather than offensive, concept. In other words, the burden of production is what a party must produce in order to avoid losing the case. Occasionally, commentators will use the term “burden of production” to refer to the showing a party must make in order to win a case at summary judgment. See, e.g., Samuel Issacharoff, Civil Procedure 56-58 (3d ed. 2012) (discussing a defendant’s initial burden in moving for summary judgment). This practice can probably be traced to the now-obsolete Supreme Court opinion in Adickes v. S. H. Kress & Co., 398 U.S. 144 (1970), in which the Court held that a defendant moving for summary judgment “had the burden of showing the absence of a genuine issue as to any material fact, and for these purposes the material it lodged must be viewed in the light most favorable to the opposing party.” Id. at 157. In Celotex Corp. v. Catrett, 477 U.S. 317 (1986), the Supreme Court essentially negated any initial “burden” on a defendant moving for summary judgment. See id. at 323 (describing the “burden” on a defendant moving for summary judgment as simply “informing the district court of the basis for its motion”). Using the term “burden of production” as a defensive concept avoids the “incomplete” and “misleading” burden-shifting discussion that can ensue when the term is used as both an offensive and defensive concept. See 21B Wright & Graham, supra note 28, § 5122, at 396-402.

\(^{52}\) See 21B Wright & Graham, supra note 28, § 5122, at 412 (“The burden of production is in the hands of the judge; the jury enforces the burden of persuasion.”).

\(^{53}\) See id. (“The sanction for failure to carry the burden of production is a [court-ordered judgment] for the adversary . . . .”)

\(^{54}\) Fed. R. Civ. P. 56(a).
would not have a legally sufficient evidentiary basis to find for the party on that issue.” In considering the burden of production, then, it is useful to recognize there are two procedural devices by which a court can preclude a fact issue from going to the jury. Although the term “burden of production” cannot be found in the Federal Rules of Civil Procedure, the two procedural devices that allow a judge to take a fact issue away from the jury are described in the Rules. In this sense, the term “burden of production” can be redefined as simply the duty of a party to assemble an evidentiary record that satisfies the requirements of Rule 56 and Rule 50 of the Federal Rules of Civil Procedure.

At first blush, the requirements of Rule 56 and Rule 50 appear to be different, as they use different language in describing the standard by which a judge determines whether a final judgment, in the absence of a jury determination, is appropriate. The reasonable jury standard has long governed the determination of whether to enter judgment as a matter of law. Before 1991, Rule 50 was silent with regard to the standard for determining when judgment as a matter of law was appropriate, but judges were relatively uniform in applying the “reasonable jury” test. In 1991, the Rule was amended to reflect this common practice, and the current language citing a “reasonable jury” was added to the Rule.

The current language in Rule 56 governing summary judgment—“no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law”—has existed in roughly the same format since the Federal Rules were adopted in 1938. And, in the 1986 case of Anderson v. Liberty Lobby,

59 See Fed. R. Civ. P. 50 advisory committee notes on 1991 amendment (describing the long use of the reasonable jury test).
60 See id. In the 1991 Amendments, the old term “directed verdict” was replaced with the modern “judgment as a matter of law.” See id.
the Supreme Court equated the standard spelled out in Rule 56 with the reasonable jury analysis used to determine whether judgment was appropriate under Rule 50. The Court noted that “[t]he primary difference between the two motions is procedural; summary judgment motions are usually made before trial and decided on documentary evidence, while directed verdict motions are made at trial and decided on the evidence that has been admitted.”

Thus, after the Supreme Court’s decision in Anderson, the term “burden of production” was synonymous with the reasonable jury standard. Simply stated, the term “burden of production” refers to the showing a party must make in order to reach the jury on a factual question. There are two mechanisms by which a federal court can take a factual issue away from the jury, and both of these Rules are governed by the reasonable jury inquiry.

As with the burden of persuasion, however, simply identifying the entity (judge or jury) with which the burden of production is concerned reveals nothing about the analysis used in determining whether a party has satisfied that burden. With regard to the burden of production, a judge conducts two separate

---

64 See id. at 251-52.
65 Id. at 251 (quoting Bill Johnson’s Rests., Inc. v. NLRB, 461 U.S. 731, 745 n.11 (1983)). The Court’s decision in Anderson equating the standard for summary judgment under Rule 56 with the standard for a judgment as a matter of law under Rule 50 is problematic when considered against the text of these two rules. The text of Rule 50 includes the “reasonable jury” language, but the text of Rule 56 does not. This issue is worth further consideration, but extends beyond the scope of this Article. For purposes of this Article, however, the reasonable jury standard will be considered only as it applies to the context of summary judgment.
67 As commonly used, the term “burden of production” differs from Rule 56 and Rule 50 only in the perspective by which the issue is addressed. The language of Rule 56 and Rule 50 considers the question primarily from the perspective of a trial court judge faced with a motion to end the case. The term “burden of production” addresses that same issue, but rather from the perspective of a litigant faced with an opposing party’s motion to win the litigation on summary judgment. The “opponent” of the summary judgment motion or judgment as a matter of law motion must satisfy her burden of production, which means that the opponent of the motion must ensure that the evidence on record demonstrates, contrary to the argument of the movant, the existence of a genuine dispute about every material fact. Simply put, the Rules naturally address the issue from the perspective of a trial judge while the term “burden of production” addresses the issue from the perspective of a party trying to reach a jury.
analyses to determine whether a plaintiff is entitled to a jury determination. First, a judge assessing whether a plaintiff has satisfied the burden of production will consider if there is evidence in the record which is sufficiently probative to justify a conclusion by the jury that the respective probability requirement has been satisfied. This aspect of the burden of production will be explored in Part II. Second, a judge measuring the burden of production will also determine whether there is sufficient record evidence such that there can be a minimum level of confidence in the probability assessment of that evidence. This aspect of the burden of production will be explored in Part III.

II. PROBABILITY AND THE BURDEN OF PRODUCTION

Ever since the Supreme Court’s decision in Anderson v. Liberty Lobby, Inc. to interpret Rule 56 according to the “reasonable jury” standard, it is clear that a judge’s analysis of the burden of production requires a judge to engage in a probability assessment. The reasonable jury standard defines a judge’s analysis of the burden of production by reference to the jury’s analysis of the burden of persuasion: A judge determining whether a plaintiff has met her burden of production must consider whether a jury would be reasonable in concluding that a

---


69 Although commentators generally do not use the phrase “probability” in discussing Anderson, there is a clear consensus that the standard adopted in Anderson permits judges to “weigh” the evidence by considering its probative value. See, e.g., Issacharoff & Loewenstein, supra note 5, at 85-86; Jeffrey W. Stempel, A Distorted Mirror: The Supreme Court’s Shimmering View of Summary Judgment, Directed Verdict, and the Adjudication Process, 49 OHIO ST. L.J. 95 (1988) (both discussing the weighing of evidence required by the Court’s decision in Anderson); see also James Joseph Duane, The Four Greatest Myths About Summary Judgment, 52 WASH. & LEE L. REV. 1523, 1561-62 (1995) (guaranteeing that some weighing of the evidence at summary judgment will be confirmed by the Supreme Court, and offering to back up that guarantee with printing costs to those who cite his article to the Supreme Court).
plaintiff has met her burden of persuasion. As discussed above, a jury’s analysis of the burden of persuasion undoubtedly requires a probability analysis. Thus, the reasonable jury standard, by directing a judge to the jury’s analysis, necessarily assumes that a judge must also consider the probability question.

While the Court’s adoption of the reasonable jury standard clearly requires a judge to engage in a probability analysis, the propriety of a judge engaging in this analysis has been questioned. The reasonable jury standard constitutes a rejection of other historical tests such as the “scintilla of evidence” and “slightest doubt” standards that seemed to preclude a judge from entering judgment against a party so long as some possible basis exists, however remote, for that party’s version of the disputed material facts.

The objections to the judge’s ability to engage in a probability analysis under the reasonable jury interpretation of Rule 56 are interesting and worthwhile. This Article, however, focuses on a different consequence of the reasonable jury test. Namely, that the reasonable jury standards masks an entirely different analysis—the confidence analysis—that judges routinely use when determining whether a party has met her burden of production.

Before turning to that topic, however, it will be useful to briefly consider the relationship, under the reasonable jury test, between the judge’s analysis of probability at the burden of production stage and the jury’s analysis of probability at the burden of persuasion stage. The frequent misunderstanding of

70 See Ronald J. Allen, How Presumptions Should Be Allocated: Burdens of Proof, Uncertainty, and Ambiguity in Modern Legal Discourse, 17 HARV. J.L. & PUB. POL’Y 627, 634 (1994) (“[B]urdens of production are a function of burdens of persuasion. A burden of production is satisfied when a reasonable person applying the relevant burden of persuasion could find in favor of the person bearing the burden of persuasion.”) (footnote omitted).

71 See, e.g., Suja A. Thomas, The Fallacy of Dispositive Procedure, 50 B.C. L. REV. 759, 775 (2009) (“When judges decide cases based on their own views of the facts—as I have shown they do under the mantra of the reasonable jury standard—judges do indeed invade the province of the jury under the Seventh Amendment, which requires in the context of dispositive motions that judges decide law and juries decide facts.”).

72 See generally Sward, supra note 6, at 575-633 (tracing the development of the reasonable jury standard as the predominate standard for measuring summary judgment, directed verdicts, and judgments notwithstanding the verdict).
this relationship might derive from efforts to depict this relationship.

The most common depiction of the jury’s probability analysis within the burden of persuasion is relatively straightforward and, by itself, uncontroversial. In this depiction, the probability of a material fact is charted on a horizontal axis, with a 0% probability on the far left and a 100% probability on the far right. The various burden of persuasion standards are then charted on the horizontal graph. The preponderance of the evidence standard is located in the middle and is usually marked with a 50% value. The clear and convincing evidence standard is to the right of the preponderance of the evidence standard, with the beyond a reasonable doubt standard to the right of the clear and convincing evidence standard. Consistent with the discussion above, the latter two standards are not identified with an exact number. Thus, the usual depiction of the burden of persuasion looks something like this:

![Figure A](image)

Problems arise, however, when commentators try to depict the judge’s probability analysis (as part of the burden of production) on the same chart as the jury’s probability analysis. In the most common variation of this effort, the plaintiff’s burden of production is charted at some point on the left of the horizontal axis; the defendant’s burden of production is charted at the right side of the chart; and the beyond a reasonable doubt measure of probability, applied in criminal law, is removed so as to make the
chart applicable to civil litigation only.73 Similar to the clear and convincing evidence measure of probability within the burden of persuasion, no exact numerical value is assigned to either the plaintiff’s or defendant’s burden of production. The clear impression from the common depiction, however, is that the probability analysis within the burden of production is a replicate of the probability analysis within the burden of persuasion, with the only difference being the specific probability associated with each concept. An example of this common depiction follows below:

![Figure B](chart.png)

**Figure B**

Because the reasonable jury test requires a judge to consider the probative value of record evidence in a manner similar to how a jury must consider the probative value of evidence introduced at trial, it is tempting to believe that Figure B, graphing both the burden of persuasion and the burden of production, is correct. However, as recognized over fifty years ago in an astute (but since forgotten or overlooked) Harvard Law Review article by the late Professor John McNaughton,74 there are fundamental analytical deficiencies in the traditional chart depicting the relationship, under the reasonable jury standard, between the judge’s probability analysis within the burden of persuasion and the jury’s

---

73 The original version of this chart appears to be attributable to legendary evidence scholar John Henry Wigmore. See 9 Wigmore, supra note 31, § 2487, at 298. Modern versions are ubiquitous. See, e.g., 21B Wright & Graham, supra note 28, § 5122, at 416.

probability analysis within the burden of persuasion. Unfortunately, Professor McNaughton’s insights have been ignored, and the traditional chart continues to be used and continues to mislead.

The first problem with the traditional chart is that it suggests that a party meets her burden of production by convincing the trial judge that her version of the disputed material fact is remotely probable. To demonstrate, consider a case in which the plaintiff’s evidence supports, at most, a 45% probability that the plaintiff’s version of the disputed material fact is true. Suppose the case is a medical malpractice suit against a doctor, who failed to perform a standard emergency operation on a patient experiencing trauma. The patient died, and the decedent’s estate brought the lawsuit. It is not disputed the doctor’s failure to operate constitutes medical negligence; the only question in the case is a factual one—whether the patient would have died even with the operation. Under the usual preponderance of the evidence measure of the probability analysis within the burden of persuasion, the plaintiff must show that, more likely than not, the plaintiff would have survived if the procedure had been performed. Suppose that the plaintiff relies exclusively on an expert’s testimony that, had the procedure taken place, the plaintiff would have had a 45% chance of living. The defendant brings a motion for summary judgment, arguing that the plaintiff has not satisfied her burden of production and, accordingly, that the judge should dismiss the case.

In this hypothetical, the judge should grant the defendant’s motion for summary judgment. Even if the plaintiff’s evidence is believed, a jury could not reasonably conclude that, more likely than not, the deceased would have survived if the operation had been performed. After all, the plaintiff’s own expert is unwilling to make this claim. Thus, the plaintiff has not met her burden to

---

75 See id. at 1384-88.
76 This assumes, of course, that there is no reframing of the plaintiff’s injury, such that the plaintiff’s injury is not death, but rather the decrease in the odds of the plaintiff surviving. See David A. Fischer, Tort Recovery for Loss of a Chance, 36 WAKE FOREST L. REV. 605, 641 (2001) (explaining the re-characterization of a plaintiff’s injury as a way to permit recovery when the usual application of the preponderance of the evidence standard to the “but-for” test for cause in fact would preclude recovery).
show that a reasonable jury could find for her on this disputed question of material fact.

Notice, however, what happens when the above hypothetical is placed on Figure C. If the plaintiff’s evidence were placed on the chart, the evidence would fall to the right of the line marking the plaintiff’s burden of production. The impression, then, is that the plaintiff in our hypothetical has cleared her burden of production and should be allowed to proceed to trial:

![Figure C](image)

The traditional chart is incorrect because it suggests that the judge in our medical malpractice hypothetical should refuse summary judgment for the defendant. As stated by Professor McNaughton, the traditional chart

implies that the burden of production is satisfied when the existence (or nonexistence) of the fact appears to be 20, 30, or some other small per cent probable. That is not so. The 50 per cent point, not the 20 or 30 per cent point, is the crucial one in the normal civil case.77

Another problem with the traditional chart is that it misrepresents the role of a judge in applying the reasonable jury analysis. The chart presumes that a juror, after hearing all of the evidence at trial, is to assign a particular probability to a material fact. So, if the material fact is whether the defendant ran a red light, the chart suggests that a jury must assign a particular probability to the likelihood of this fact being true. According to

77 McNaughton, supra note 74, at 1385 (footnote omitted).
the chart, if the juror assigns a greater-than-fifty-percent chance (and assuming the typical preponderance of the evidence measure of probability), the juror must consider the material fact as being proved by the plaintiff.

The traditional chart fails to capture the essence of the reasonable jury standard because the chart suggests that a judge should also assign a specific probability value to a material fact in applying the reasonable jury standard. As the Supreme Court has explained, however, under the reasonable jury standard the trial court judge is not to "determine the truth of the matter but to determine whether there is a genuine issue for trial." So, in evaluating the burden of production, a judge should not precisely determine the probability that she would assign to the material fact, but rather assign a range of probabilities that form a reasonable conclusion from the evidentiary record.

Similar to the first problem with the traditional chart, this second problem was also identified by Professor McNaughton long ago:

It is more difficult to plot the burden of production of evidence than it is to plot the burden of persuasion, because the burden of production involves not just a determination of the probabilities of a fact by a jury. It involves also an estimate by the judge as to the limits within which such determinations might fall. The judge in whose mind the "reasonable jury" exists is a reasonable judge, and he has had experience with many real juries; but what, he must ask, would a "reasonable jury" believe the probabilities to be? What is the maximum reasonable probability? What is the minimum reasonable probability? The judge's estimate therefore is not a single point—a single probability value. It is a range of probability values falling between two points.

Professor McNaughton offered a depiction (reproduced in Figure D, below) that resolves the difficulties he highlighted. Professor McNaughton's diagram charts the plaintiff's burden of

---

79 Under mathematical terminology, this is known as a "point estimate." See infra text accompanying note 167.
80 McNaughton, supra note 74, at 1386.
81 See id. at 1387 (explaining how his depiction resolves the difficulties he noted).
persuasion identically to the traditional chart. The probability analysis within the burden of production, however, is not a single point on a line but rather the range of probabilities that a judge believes are a reasonable conclusion as to a material fact from the record evidence.

The depiction above represents a case in which a judge believes that a reasonable juror could find a material fact (required to be proven by the plaintiff) either more than 50% or less than 50% probable. The gray shaded box represents this analysis. In this hypothetical case, the burden of production is met for both the plaintiff (who must prove the material fact at trial) and the defendant (who can win at trial by showing the material fact does not exist) because the judge believes that a reasonable jury could come to a conclusion for either the plaintiff or the defendant. Because the hypothetical case at issue involves the typical preponderance of the evidence measure of probability, the case must proceed to trial so long as the gray box straddles the important 50% mark.

However, when the range of a jury’s reasonable assessment of probability falls exclusively on either side of the 50% mark, summary judgment might be appropriate.82 Thus, in Figure E, the defendant is entitled to summary judgment because a reasonable

---

82 Or, in the case of a defendant’s motion for summary judgment, a range of probabilities that falls between exactly 50% and higher would also warrant summary judgment because of the well-established notion that the defendant gets the benefit of a “tie.” See supra note 44 and accompanying text.
jury could reach only one conclusion as to the material fact that the plaintiff needs to prove as part of her prima facie case:

![Figure E](image)

**Figure E**

In Figure F, conversely, the probabilities assigned by a reasonable jury exceed the plaintiff's burden of persuasion; thus, there is no genuine dispute regarding this material fact:

![Figure F](image)

**Figure F**

In Figure F, the plaintiff has “won” this issue and might be entitled to a judgment on her behalf, depending on whether there are genuine disputes about other material facts.83

83 In the second chart, although a reasonable jury must conclude that the plaintiff has met her burden of persuasion with regard to this material fact, summary judgment for the plaintiff is not necessarily appropriate if there are other facts in the litigation involving a genuine dispute. A defendant can win summary judgment by showing that the plaintiff has failed to meet her burden of production on one material fact. If a reasonable jury would find the material fact in favor of the defendant, there is no need to proceed to trial because the defendant cannot be found liable. However, for a plaintiff to win a summary judgment, the defendant must fail to meet his burden of production for every material fact. If there is any genuine dispute about any material fact, summary judgment for the plaintiff is not appropriate. This concept, intuitive to all judges and lawyers, is not necessarily evident from the language of Rule 56, which states that a judge “shall grant summary judgment if the movant shows that there is no genuine dispute as to any material fact.” Fed. R. Civ. P. 56(a).
Professor McNaughton’s more accurate depictions of the burden of persuasion and the reasonable jury analysis of the burden of production confirm the Supreme Court’s decision in *Anderson v. Liberty Lobby, Inc.* that, in evaluating whether a plaintiff has satisfied the probability analysis within the burden of *production*, the measure of probability applicable to the jury’s resolution of the burden of *persuasion* must be considered. In *Anderson*, the plaintiffs brought a libel claim against defendants. Because plaintiffs were “public figures,” the First Amendment required that plaintiffs prove the defendants’ “actual malice” under the clear and convincing evidence standard. The trial court granted the defendants’ motion for summary judgment. The plaintiffs appealed to the United States Court of Appeals for the District of Columbia Circuit, which partially reversed the trial court’s grant of summary judgment. In reversing, the D.C. Circuit reasoned that the higher measure of probability at trial was irrelevant for purposes of determining whether the defendants were entitled to summary judgment. The Supreme Court reversed the D.C. Circuit on this point and the logic of the Court’s conclusion is evident from the charts used in this part to depict the probability analysis within the burdens of production and persuasion. Under the reasonable jury standard, the critical question for a trial court judge is whether a reasonable jury could assign probabilities on either side of the probability threshold required by the applicable measure of probability at the burden of persuasion stage. The measure of probability at trial is critical to determining whether the measure of probability at the burden of production stage has been met. Obviously, altering the measure of probability can affect whether the judge-assigned reasonable jury range of probability extends across the critical threshold. If the measure of probability at trial is moved, this can

---

85 *See id.* at 245.
86 *Id.* at 254.
87 *Id.* at 252.
88 *See id.* at 246.
89 *See id.*
90 *See id.*
91 *See id.* at 246-47.
92 *See id.* at 244.
affect whether the reasonable jury range of probability assigned by the judge extends across this critical point. Accordingly, in Figure G below, a trial judge would have to determine that the plaintiff’s burden of production, with regard to the probability question, was met under the preponderance of the evidence measure of the burden of persuasion (and thus deny a defendant’s motion for summary judgment), but not under the clear and convincing evidence measure of the burden of persuasion (and thus grant a defendant’s motion for summary judgment).

Figure G

A final point regarding the difference between direct and circumstantial evidence will complete the discussion in this section on the judge’s probability analysis under the burden of production. Every litigator (and perhaps every attorney) is familiar with the notion that weighing the credibility of witnesses is the exclusive task of the jury.\(^{93}\) This principle limits the application of the judge’s analysis of probability within the burden of production. When a party presents direct evidence on a material fact, the judge is required to assign a probability range that satisfies that party’s burden of production under the reasonable jury inquiry.\(^{94}\) Thus, if the plaintiff must prove a material fact and has direct testimony from a witness that a material fact occurred, the range of probabilities assigned by the judge must extend, to

---

\(^{93}\) See Miguel A. Méndez, Witnesses: Conforming the California Evidence Code to the Federal Rules of Evidence, 39 U.S.F. L. Rev. 455, 465 (2005) (“Trial lawyers know that the outcome of a trial will be determined in almost all cases by the witnesses the jurors choose to believe and the ones they decide to ignore.”).

\(^{94}\) See Duane, supra note 69, at 1574-80 (describing the rule that a trial court judge is not to weigh the credibility of witnesses as “the prevailing view embraced by virtually every commentator and now accepted by every court of appeals to squarely address the issue,” but then suggesting that this rule is incorrect).
the right, past the relevant measure of probability at the burden of persuasion stage. On the other hand, if the defendant produces direct testimony that the material fact did not occur, the range of probabilities assigned by the judge must extend, to the left, past the burden of persuasion point. Further, when both parties submit direct evidence on a material fact, the judge must assign a range of probabilities that extends across the relevant measure of probability. Obviously, when both parties produce direct evidence on a material fact, both will meet the probability component of the burden of production. The reasonable jury analysis only requires a judge to establish a range of probabilities involving a material fact when one—or both—parties seek to prove a material fact through circumstantial evidence alone.

III. CONFIDENCE AND THE BURDEN OF PRODUCTION

The “reasonable jury” standard works fairly well in describing the relationship between judge and jury with regard to the probability question. Despite the problems, discussed above, that sometimes arise in depicting or discussing the reasonable jury standard, once these problems are resolved the reasonable jury standard is descriptively accurate in portraying the relationship between judge and jury with regard to probability.

Unfortunately, the reasonable jury standard has the side effect of masking a completely different type of analysis within the burden of production. In reality, judges routinely grant summary judgments according to an analysis that is distinct from a probability analysis. This different type of analysis will be labeled as a “confidence” analysis. To introduce this “confidence” concept—and also to take advantage of several insights already made by academics on this topic—two famous hypotheticals will be considered: the “gatecrasher” and “blue bus” hypotheticals. These two hypotheticals provide a convenient cornerstone in

---

95 See William Daniel Quarles et al., Summary Adjudication: Dispositive Motions and Summary Trials § 5.14, at 118 (1991) (“When judges are confronted with conflicting assertions of historical fact, for example, plaintiff’s witness asserts that the light was red, while defendant’s witness is equally adamant that the light was green, there is an issue of material fact that may not properly be resolved on summary judgment.”).
building the argument that a probability analysis is not the exclusive standard for measuring a party’s burden of production.

A. The Blue Bus and Gatecrasher Hypotheticals: Descriptive Questions

The famous “gatecrasher” hypothetical was first posed by Professor L. Jonathan Cohen in his book “The Probable and the Provable.”

Consider, for example, a case in which it is common ground that 499 people paid for admission to a rodeo, and that 1,000 are counted on the seats, of whom A is one. Suppose no tickets were issued and there can be no testimony as to whether A paid for admission or climbed over the fence. So by any plausible criterion of mathematical probability there is a .501 probability, on the admitted facts, that he did not pay. The mathematical theory would apparently imply that in such circumstances the rodeo organizers are entitled to judgment against A for the admission-money, since the balance of probability (and also the difference between prior and posterior probabilities) would lie in their favour. But it seems manifestly unjust that A should lose his case when there is an agreed mathematical probability of as high as .499 that he in fact paid for admission.

Indeed, if the organizers were really entitled to judgment against A, they would presumably be equally entitled to judgment against each person in the same situation as A. So they might conceivably be entitled to recover 1,000 admission-moneys, when it was admitted that 499 had actually been paid. The absurd injustice of this suffices to show that there is something wrong somewhere. But where?96

The “gatecrasher” hypothetical is rivaled in fame (or infamy) by Professor Laurence Tribe’s97 “blue bus” hypothetical. Professor Tribe’s original hypothetical is rather straightforward and simple:

97 According to Professor Tribe, he did not invent this hypothetical. See Laurence H. Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 HARV. L. REV. 1329, 1341 n.37 (1971) (calling the blue bus hypothetical a “famous chestnut”). Nevertheless, the origin of this hypothetical is generally attributed to Professor Tribe. See Jonathan J. Koehler & Daniel N. Shaviro, Veridical Verdicts: Increasing Verdict
Consider next the cases in which the identity of the responsible agent is in doubt. Plaintiff is negligently run down by a blue bus. The question is whether the bus belonged to the defendant. Plaintiff is prepared to prove that defendant operates four-fifths of all the blue buses in town. What effect, if any, should such proof be given?²⁹⁸

These two relatively simple hypotheticals have bedeviled legal scholars interested in procedure and evidence and have attracted the interest of non-legal scholars from a variety of disciplines.³⁰⁰ In 1986, an excellent symposium hosted by the Boston University Law Review attracted leading scholars from various fields; the primary fodder for the exhaustive academic discussion that ensued were these two hypotheticals.³⁰¹ Although they were introduced decades ago, numerous contemporary discussions of these two hypotheticals can still be

---

²⁹⁸ Tribe, supra note 97, at 1340-41.

²⁹⁹ See, e.g., James Brook, The Use of Statistical Evidence of Identification in Civil Litigation: Well-Worn Hypotheticals, Real Cases, and Controversy, 29 St. Louis U. L.J. 293, 299 (1985) (“The Blue Bus Company hypothetical is a classic, and we can see why: like all good hypotheticals it raises more questions than it answers.”).


³⁰² For the uninitiated in formal mathematical theory, one might add “exhausting” to the list of adjectives describing the work product from this symposium. Cf. 21B Wright & Graham, supra note 28, § 5122, at 410 (“Some writers have tried to puff up the preponderance standard with academic hot air, including fancy statistical analysis.”).

³⁰³ See Green, supra note 101, at 378 (acknowledging the importance of these two hypotheticals to the topic of the symposium).
found in legal literature, particularly commentary, devoted to issues of criminal law.

To grasp the array of points commentators make with regard to the gatecrasher and blue bus hypotheticals, the descriptive question and the normative question implicated by these hypotheticals must be differentiated. The descriptive question examines how real courts would handle these hypotheticals if they represented real fact patterns. The normative question examines how these hypotheticals should be handled.

When commentators address the descriptive question, they disagree on it. Most presume that a plaintiff relying only upon this statistical evidence would fail to meet the burden of production.

---


106 See Brook, supra note 99, at 299 (describing how academics have devoted most of their effort to discussing the normative, rather than descriptive, issues raised by the gatecrasher and blue bus hypotheticals, and also noting that when the descriptive question is addressed one might be surprised with “the ease with which most writers seem able to satisfy themselves regarding the answer”).

107 The commentators who have considered the burden of production question raised by the blue bus and gatecrasher hypotheticals have usually done so in the context of a directed verdict. See, e.g., Richard Lempert, The New Evidence Scholarship: Analyzing the Process of Proof, 66 B.U. L. REV. 439, 457-58 (1986) (discussing the gatecrasher hypothetical in the context of a directed verdict). Commentators have generally ignored whether their conclusions applied to summary judgment as well. See William Twining, The Boston Symposium: A Comment, 66 B.U. L. REV 391, 397 (1986) (noting the absence of any discussion as to how the gatecrasher and blue bus hypotheticals affected pre-trial procedure). Of course, at the time the blue bus and gatecrasher hypotheticals were receiving the most scholarly attention, summary judgment was only in the beginning process of being transformed by the summary judgment trilogy of cases. In the nearly five hundred pages of articles published in the 1986 Boston University Law Review Symposium, the important trilogy of cases were barely recognized; the Anderson v. Liberty Lobby, Inc. case was cited
Professor Lea Brilmayer states confidently: “Of course, the law would not allow recovery in either of these cases.” Similarly, Professor Charles Nesson concludes: “In [the blue bus] case and others like it, the plaintiff will lose; in fact, the case is unlikely even to reach the jury.” The creators of the gatecrasher and blue bus hypotheticals, Professor Cohen and Professor Tribe, also seem to presume that a plaintiff would lose under these hypotheticals. Others, however, suggest that a plaintiff could take this case to a jury and that a judgment in the plaintiff’s favor would not be disturbed. Still others take the more cautious (but—most likely—more accurate) view that it is unclear how courts would respond when confronted with these exact hypotheticals, and courts might not give a uniform response to these fact patterns.


Nesson, supra note 22, at 1379; see also id. n.70 (citing cases).

See Cohen, supra note 96, at 75-76 (discussing how a victory for the plaintiff seems to be required by the preponderance of the evidence standard, but implying that this “absurd injustice” would not actually occur in a real case); Tribe, supra note 97, at 1349 (“To return once again to the blue bus litigation, even assuming a standard of proof under which the plaintiff need only establish his case ‘by a preponderance of the evidence’ in order to succeed, the plaintiff does not discharge that burden by showing simply that four-fifths, or indeed ninety-nine percent, of all blue buses belong to the defendant.”) (footnote omitted).

See Richard D. Friedman, Generalized Inferences, Individual Merits, and Jury Discretion, 66 B.U. L. REV. 509, 514 (1986) (“[Professor] Zuckerman and others assume out of hand that the bus and rodeo hypotheticals clearly would not reach the jury. But at least one case, Kaminsky v. Hertz Corp., suggests otherwise.”) (footnotes omitted); Fienberg, supra note 100, at 698 (“At least two sitting federal judges have told me that they would clearly allow the case to go to the jury and that the only argument for overriding such evidence is one from spoliation.”); Lempert, supra note 107, at 461 (“In these circumstances [such as those presumed in the gatecrasher and blue bus hypotheticals], where the plaintiff cannot present non-statistical evidence and we cannot refer to other values, I think the intuitions that the plaintiff should not recover and that the law would not allow recovery are incorrect.”).

See Brook, supra note 99, at 316 (“As I have tried to demonstrate in the previous section, the case law in the area is not of much help. It is far from clear what the court would do if the facts were well presented and, after various complicating factors were thoroughly examined, the quantified evidence still weighed against the defendant.”) (footnote omitted).
the exact parameters of the gatecrasher and blue bus hypotheticals. Commentators frequently rely on *Smith v. Rapid Transit, Inc.* for the notion that the blue bus and gatecrasher hypotheticals would end in a court-directed judgment for the defendant. In *Smith*, which was the basis for the blue bus hypothetical, the Massachusetts Supreme Court affirmed a directed verdict for the defendant, who operated a bus company. The plaintiff had been injured while avoiding an oncoming bus on Main Street, and the defendant had the sole franchise for operating a bus route on this street. From the Massachusetts Supreme Court’s opinion, it is apparent that the plaintiff presented some direct testimony as to the description of the bus involved in the accident during her case-in-chief. Apparently, however, this testimony was not very probative on the question of whether defendant’s bus was the one involved in the accident, because the Massachusetts Supreme Court focused solely on the fact that defendant had the exclusive franchise for operating a bus line on Main Street. According to the court, this evidence was not sufficient for plaintiff to meet her burden of production and avoid a directed verdict: “The 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.”

The *Smith* opinion does provide some support for the argument that the gatecrasher and blue bus hypotheticals would end in a judge-directed judgment for the defendant, but the case does not exactly replicate the hypotheticals. Although both the actual *Smith* case and the two hypotheticals involve the plaintiff’s reliance on circumstantial evidence to prove a material fact (identity), the hypotheticals involve an exact statistical

---

113 58 N.E.2d 754 (Mass. 1945).
115 See *Smith*, 58 N.E.2d at 755.
116 See id. at 754.
117 See id. at 754-55.
118 See id. at 755.
119 See id. at 754.
120 See id. at 755.
121 See id. at 755.
122 Id.
probability. Thus, from the assumed evidence in the blue bus hypothetical, there is an 80% probability that the defendant owned the bus in question, while in the gatecrasher hypothetical there is a 50.1% chance that the defendant was a gatecrasher. In the actual Smith case, however, no such exact probability is produced, despite the court’s acknowledgement that the “mathematical chances somewhat favor the proposition that a bus of the defendant caused the accident.”\textsuperscript{122} The Smith court’s statement appears to be based on a rough estimate of the frequency that other buses, not owned by the defendant’s company, traveled down Main Street, as opposed to the mathematical precision of the probability inference that could be drawn from the circumstantial evidence in the blue bus and gatecrasher hypotheticals. For those who claim Smith as support for their position that the blue bus and gatecrashers hypotheticals would end in a directed verdict, this difference might be relevant, depending on how they explain the result in the Smith case and the hypotheticals.

There are other differences between Smith and the hypotheticals, too. The gatecrasher and blue bus hypotheticals are based on the assumption that there is simply no other evidence on the identity question and that neither party is hiding more specific evidence that might shed light on the material fact.\textsuperscript{123} In other words, the assumption is that the statistical evidence is the only available evidence. This assumption does not necessarily apply to the Smith case. In the Massachusetts Supreme Court’s opinion, the court identifies testimony from the plaintiff describing the bus in question, so it seems unlikely that the evidence involving the bus route, to which the court assigned a “mathematical chance[]”\textsuperscript{124} favoring the plaintiff, was the only evidence available as to the identity of the bus. Another difference that could be important for the specific thesis of this paper, which focuses on federal law, is that the Smith case occurred in Massachusetts state court. This raises the possibility that Massachusetts’s procedural law and discovery law were

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{122} Id.
\item \textsuperscript{123} See Lempert, supra note 107, at 458 (explaining that with hypotheticals “the impossibility of any more specific identification is clear”).
\item \textsuperscript{124} See Smith, 58 N.E.2d at 755.
\end{itemize}
\end{footnotesize}
sufficiently unique (in 1945) so as to complicate attempts to draw extensive modern conclusions regarding federal law from *Smith*.

Other cases cited by academics to support their descriptive arguments involving the blue bus and gatecrashers hypotheticals suffer from the same deficiencies as the *Smith* case; in some way or another, these actual cases differ materially from the controlled fact pattern of the hypotheticals. Upon reflection, this difference is perhaps not surprising. The usual benefit of a hypothetical is that it can isolate a difficult legal issue with a precision that does not exist in actual litigation. The blue bus and gatecrasher hypotheticals accomplish this objective, but the price of this precision is that the hypotheticals become somewhat removed from the circumstances of typical cases.

A conclusive resolution of the descriptive question raised by the blue bus and gatecrasher hypotheticals is not critical to the thesis of this paper. The most accurate premise is probably that different courts would react differently to the two hypotheticals. For the purposes of this paper, it is only necessary to establish that *some* courts would resolve the hypotheticals by entering a judgment for the defendant—as a matter of law—without the case reaching the jury. Most academics have reached this conclusion when addressing this question, and the *Smith* case seems to support this limited proposition.

**B. The Blue Bus and Gatecrasher Hypotheticals: Normative Explanations**

Assuming that at least some courts would conclude, in the blue bus and gatecrasher hypotheticals, that the plaintiff had not met her burden of production and would thus enter a judgment for the defendant, the next question is normative: Would a court be correct in doing so, and, if so, how can this result be explained? The academic perspectives on the normative question are more unified in the sense that, with a couple notable exceptions, commentators have generally agreed that a court would be correct

---


126 See *supra* notes 106-10 and accompanying text.
in entering judgment for the defendant. Despite this general agreement, commentators take a variety of approaches in explaining the “paradox” of the gatecrasher and blue bus hypotheticals. The paradox of the gatecrasher and blue bus hypotheticals requires an explanation as to why a court should enter judgment against the plaintiff when the only available evidence suggests that the material fact, necessary for the plaintiff’s recovery, is more probable than not? In other words, how can a court enter judgment on behalf of a defendant when a probability analysis of the available evidence seems to suggest only one reasonable conclusion from the evidence—a conclusion in favor of the plaintiff rather than the defendant?

The most dramatic explanations for this paradox are those offered by the creators of the hypotheticals. According to Professor Cohen, the example of the gatecrasher is one of many paradoxes within the law that necessitates a rejection of the traditional—or, as Cohen terms it, Pascalian—understanding of probabilities. One of the primary tenets of the traditional, or Pascalian, view of probabilities is that the probability of “A” and the probability of “not A” must always equal 100%. Professor Cohen believes that his gatecrasher hypothetical undermines the validity of this tenet of traditional probability because it allows a plaintiff to recover when the mathematical probability of a fact material to the plaintiff’s case has a probability of 50.1%, meaning that the

127 See Brook, supra note 99, at 321 (“I have found that many people, including those well versed in legal theory and whose opinions I would not want to quickly dismiss as unthinking or illogical, react in much the same way to cases like those against the accused gatecrasher or the Blue Bus Company. There is for many a strong feeling that finding against any defendant on such evidence alone is somehow improper.”) (footnote omitted).

128 See COHEN, supra note 96, at 2-3 (using “Pascalian” as term representing conventional mathematical understanding of probability).

129 See Twining, supra note 107, at 392 (describing Cohen’s rejection of Pascalian logic in favor of Baconian logic).

130 See COHEN, supra note 96, at 51 (“Again, the mathematical probability of an event obeys a complementational negation principle whereby the probability of the event’s occurrence and the probability of its non-occurrence always sum to 1.”). This concept is implicit in “Bayesian statistics,” which require an assessment of the probability and non-probability of an event to equal one. See Bruce Abramson, Blue Smoke or Science? The Challenge of Assessing Expertise Offered as Advocacy, 22 WHITTIER L. REV. 723, 768-72 & n.156 (2001) (explaining Bayesian reasoning and also recognizing the necessity that all probabilities continue to “sum to 1.0”).
defendant is exposed to liability even though there is a 49.9% chance that the material fact (the defendant’s “crashing” of the gate without paying) did not occur. Professor Cohen attempts to rectify the gatekeeper paradox, as well as other problems identified as being inconsistent with the Pascalian understanding of probabilities, with a new understanding of probabilities that

131 The irony with Professor Cohen’s gatekeeper example is that he seems to be using the example to demonstrate one problem that most lawyers would not find particularly difficult, and in the process Cohen happened to isolate a problem that jurists do find troubling. As alluded to in the text, Professor Cohen seems to believe that the problem with the gatecrasher hypothetical is that the defendant would be found liable—if normal rules were applied—even though there is a 49.9% chance that he had paid for his ticket to the rodeo. See COHEN, supra note 96, at 75. Most lawyers, judges, and law professors would not find this particular aspect of the hypothetical troubling. Lawyers, judges, and law professors are comfortable with the idea that a loss has to fall on someone and, further, can accept that hard choices have to be made in close cases and that our legal system does its best to place those losses on the appropriate party. See Richard Eggleston, The Probability Debate, 1980 Crim. L. Rev. 678, 681-82 (“[T]he injustice which troubles Mr. Cohen, of giving judgment against a man whose case has a probability of 0.499 of being true, exists in every case in which the plaintiff makes a prima facie case on the balance of probability . . . .”). Thus, most jurists would disagree with Professor Cohen’s statement that “it seems manifestly unjust that [the defendant] should lose his case when there is an agreed mathematical probability of as high as .499 that he in fact paid for admission.” COHEN, supra note 96, at 75. From a jurisprudential perspective, however, the real issue pinpointed by Cohen’s hypothetical is the method by which the 50.1% and 49.9% figures are adduced. If a jury heard direct testimony from both the plaintiff and the defendant in the gatecrasher hypothetical, and determined after listening to this testimony that there was a 50.1% chance that the defendant was a gatecrasher, most jurists would say that there is no “paradox” in holding the defendant liable. The real paradox of the gatecrasher hypothetical, at least from a jurisprudential perspective, is that the 50.1% and 49.9% figures in the gatecrasher hypothetical were derived from circumstantial (and statistical) evidence that is only a small part of the full range of evidence that might be probative on the defendant’s liability. It is not clear that Cohen ever recognized that the paradox he saw in the gatecrasher hypothetical was different than the paradox that jurists saw in the gatecrasher hypothetical. But, at least, it is clear why Cohen used his statistical circumstantial evidence in his hypothetical. The notion that a jury would assign a 50.1% probability to a material fact (as I did above) based on the evaluation of direct evidence is absurd. Thus, Cohen needed to use statistical evidence to make realistic (and to dramatize) his “paradox” that a defendant would be liable—under a standard application of the “normal” rules—even though the plaintiff had only proven the material fact to a 50.1% probability. In the process, though, Cohen opened up an entirely different can of worms.

132 Another of Cohen’s problems with the traditional understanding of probabilities is the “difficulty about conjunction.” See COHEN, supra note 96, at 59. Under this “difficulty,” if a plaintiff must prove Fact A and Fact B to recover from a defendant, a plaintiff who proves both to a 60% probability would presumably win her case under existing law. But, under the traditional mathematical, or Pascalian, view of
he labels “inductive probabilities.”¹³³ Cohen’s actual argument is somewhat dense and difficult to digest for those¹³⁴ without a sufficient mathematical and philosophical background,¹³⁵ a fact seemingly recognized by Cohen himself.¹³⁶ Nevertheless, a simplified summation of Cohen’s theory of inductive probabilities is useful because Cohen’s ideas establish an important theme (albeit from a formal, mathematical perspective) that is echoed in some of the more practical legal perspectives on the gatecrasher hypothetical. Professor Marjorie McDiarmid provides just such a summation of Cohen’s theory of inductive probabilities (and an excellent one at that):

In the closed universe of [“traditional”] mathematical probability, if one does not totally believe in the truth of a proposition P, one perforce assigns that uncertainty to a belief in the converse proposition P. Thus if one were so unwise as to say, “I only believe in leprechauns .1,” one has effectively said, “I believe in no leprechauns .9.” Cohen points out another option. One could say, “I lack sufficient information to believe in leprechauns more than .1, but for the same reason I [also] lack sufficient information to believe in no leprechauns [more than] .1 . . . . I just don’t know.” So while in mathematical probability zero (0) equals absolute disbelief in a proposition and conversely absolute belief in its negation, in

³³ Cohen, supra note 96, at 121.
³⁴ Like me and, I suspect, the typical reader of this law review article.
³⁵ See Brook, supra note 99, at 313 (“To the reader who comes to Cohen’s book because of an interest and background in law, this section [in which Cohen develops inductive probability] may be virtually impenetrable.”).
³⁶ See COHEN, supra note 96, at 4 (suggesting that those with a “primary interest . . . in jurisprudence” read the chapters of the book out of order and that the dense mathematical portions of the book could be read only as a matter of “theoretical completeness”).
Cohen’s system zero (0) means insufficient information to believe anything, either the proposition or its converse.137

By rejecting the notion that the probability of both “A” and “not A” must equal 100%, Cohen is able to identify a concept (from his mathematical perspective) that other commentators have also found important (from legal perspectives) to an understanding of the gatecrasher paradox: This concept is of the sufficiency of information from which to make a probability analysis. As will be shown below, this concept looms large in the best efforts of legal scholars to explain the gatecrasher and blue bus hypotheticals. Further, it is at the heart of the confidence principle developed in this paper.138

Like Professor Cohen, Professor Laurence Tribe uses his hypothetical (the blue bus hypothetical) as part of a discussion of a larger theme.139 For Professor Tribe, that theme is the propriety of statistical evidence.140 In his article containing the blue bus hypothetical, Professor Tribe explains his reluctance to wholeheartedly embrace statistical evidence.141 Tribe’s reluctance is based primarily on a fear that the probative value of statistical evidence might be weighted too heavily by jurors who are not comfortable with the use of statistics in decision-making.142 Whereas Professor Cohen’s mathematical rejection of the traditional understanding of probabilities established a theme—sufficiency of information—that has sometimes been reflected in the ensuing academic discussion of these two hypotheticals,


138 In light of the irony identified in note 131, there is further irony that Professor Cohen’s reconceptualization of the traditional understanding of probabilities would be of benefit to legal scholars. After all, if my observations in note 131 are correct, Professor Cohen was concerned with a completely different problem than the one that piqued the interest of legal scholars. Professor Cohen’s solution, though, appears to be broad enough so as to be of at least some benefit to both “problems.”

139 See Tribe, supra note 97.

140 See id.

141 See id. at 1331 (“[M]y central concern is the wisdom of using mathematical methods for certain decisionmaking purposes even when those methods are rationally employed . . . .”).

142 See id. at 1360-61 (explaining the “risk that the jury will give [statistical evidence] too much weight when undertaking to combine” it with other types of evidence).
Professor Tribe’s argument has similarly colored the tone of the subsequent academic discussion. Professor Tribe’s focus on statistical evidence is replicated in various academic papers discussing the blue bus and gatecrasher hypotheticals. In fact, as explored in another Article, one of the primary deficiencies in the academic discussion of the blue bus and gatecrasher hypotheticals is a failure to recognize that the principles discussed by the commentators can be applied to cases involving non-statistical (but circumstantial) evidence.

Outside of the broad themes struck by Professors Cohen and Tribe, commentators offer a variety of more specific explanations as to why the plaintiffs in both the gatecrasher and blue bus hypotheticals should lose before ever reaching a jury. Professor Adrian A.S. Zuckerman suggests that it is “unjust” to allow the plaintiff in the gatecrasher hypothetical to recover because of our moral objection to “corporate liability”:

As a principle for the imposition of responsibility, corporate punishment is characterized by the assumption that it is justified to hold an entire social group responsible for the transgressions of its individual members. Our moral and legal values strongly resist this principle because it fails to acknowledge that the individual is entitled to judgment on his own actions.

The moral objection to corporate liability is of course underlined by our personal concern not to be held responsible for the actions of another. Judgments based on naked statistical distributions openly acknowledge that the individual defendant may well belong to the innocent minority, and therefore undermine the citizen’s confidence that the legal system will protect him from mistaken conviction of crime or mistaken imposition of liability. Yet such confidence is crucial to public respect for judicial adjudication.

---

143 See Friedman, supra note 111, at 512-13 (discussing the gatecrasher and blue bus hypotheticals as they pertain to “naked statistical evidence”).
145 Adrian A.S. Zuckerman, Law, Fact or Justice?, 66 B.U. L. REV. 487, 489 (1986); see also Peter Tillers, If Wishes Were Horses: Discursive Comments on Attempts to
Professor Charles Nesson sounds a somewhat similar theme in his article “The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts.” In his article, Professor Nesson posits that it is critical that jury trials be accepted as legitimate by the general public, and that this “depends on a court’s ability to cast a verdict not as a statement about the evidence presented at trial, but as a statement about a past act—a statement about what happened.” Professor Nesson believes that this concept explains why a judge should grant a judgment for the defendant in the blue bus hypothetical:

Although the defendant probably caused the plaintiff’s injury, the factfinder cannot reach a conclusion that the public will accept as a statement about what happened. The factfinder can only conclude from the plaintiff’s evidence that there was an 80% chance that he was injured by the Blue Bus Company and a 20% chance that he was not. What is crucial in this situation is that the public cannot view whatever statement the factfinder makes as anything other than a bet based on the evidence. Given the evidence’s simplicity, the public has no reason to defer to a factfinder’s conclusion as in cases involving complex evidence. And given the evidence’s statistical nature, the public need not defer to the conclusion on the ground that the factfinder is in a better position to evaluate a witness’s demeanor. In short, the factfinder cannot, and the public knows it cannot, make anything other than a bet on the evidence. Because the judicial system strives to project an acceptable account about what happened, then, the plaintiff’s evidence is insufficient, notwithstanding the high probability of its accuracy.

---

146 Nesson, supra note 22, at 1357.
147 Id. at 1358.
148 Id. at 1379.
Professor Richard Lempert takes a slightly different tactic in explaining the gatecrasher and blue bus hypotheticals.\textsuperscript{149} Professor Lempert suggests that a judge-ordered judgment for the defendants would encourage plaintiffs to come forward with more specific evidence rather than the generic, statistical evidence considered in those hypotheticals.\textsuperscript{150} Professor Lempert’s approach has been criticized for seemingly changing the parameters of the hypotheticals by assuming that more evidence is available.\textsuperscript{151} Professor Lempert notes, though, that it will often be unclear if other evidence (besides naked statistical evidence) is available.\textsuperscript{152} Because of this uncertainty, per Professor Lempert, the judgment for the defendants in the gatecrasher and blue bus hypotheticals makes sense because it places an incentive on plaintiffs to find additional information: “If plaintiffs are allowed to get to the jury by showing that naked statistics are the only evidence available, they will have an incentive to falsely create the appearance that this is the case.”\textsuperscript{153} By directing a verdict for the defendants in cases like the blue bus and gatecrasher hypotheticals, the court discourages plaintiffs from bringing an actual case resting on pure statistical evidence absent a very good explanation for the lack of specific evidence regarding the defendant’s liability.\textsuperscript{154}

While the insights of Professors Zuckerman, Nesson, and Lempert are helpful to understanding the paradox of the blue bus and gatecrasher hypotheticals, they do not resolve the underlying root problem of how to justify the judgment for the defendant in terms of the burden of persuasion and burden of production. Under the probability analysis within the burden of persuasion, the defendant is—more likely than not—liable in each of these

\textsuperscript{149} See generally Lempert, supra note 107.

\textsuperscript{150} See id. at 460 (discussing the incentive on plaintiffs to come forward with more evidence).

\textsuperscript{151} See id. (“This brings me to the charge that I, like Kaye, have ducked the hypothetical.”); see also Allen, supra note 125, at 412 (“The only sensible way to understand the [gatecrasher] hypothetical is that it presents the question of what should be done when this is all the evidence there is.”).

\textsuperscript{152} See Lempert, supra note 107, at 457 (“[I]n the real world we will never be sure if the conditions of the hypothetical are met; much more often than not they won’t be met, and more information will be available to the plaintiff.”).

\textsuperscript{153} Id. at 460.

\textsuperscript{154} See Green, supra note 101, at 378 (commenting that a plaintiff’s lawyer would be unlikely to bring suit based only naked statistical evidence).
hypotheticals. Under the reasonable jury standard, a judge would have to conclude that any deviation from the statistical probabilities established by the circumstantial evidence was unreasonable. Thus, in these hypotheticals, there can be no “reasonable” disagreement as to the probability assessment of the evidence. A depiction of the judge’s probability analysis in the gatecrasher and blue bus hypotheticals would look like this:

![Figure H](image)

**Figure H**

Figure H does nothing to support the notion that a judge should (and would) enter a judgment for the defendant before trial. If anything, Figure H suggests the opposite: That the plaintiff, rather than the defendant, is entitled to judgment at the burden of production stage in advance of a jury trial.

Obviously, there is something more at play in the blue bus and gatecrasher hypotheticals than a probability analysis. If the burden of persuasion and burden of production are understood only in terms of probability, a judge-ordered judgment for the defendants in these two hypotheticals cannot be justified. The insights of Professors Zuckerman, Nesson, and Lempert provide reasons why the result suggested by a probability understanding is problematic, yet they do not systematically deal with how the burdens of production and persuasion should be understood to correlate to the result they (as well as most commentators) believe to be correct.

Towards this goal, Professor Jonathan Cohen’s insights are more helpful. Recall Professor Cohen’s theory of inductive probabilities, in which the probability of an event and the non-probability of an event do not have to equal 100% if one considers
the idea of sufficiency of information. The idea of sufficiency of information can explain the result in the gatekeeper and blue bus hypotheticals in terms of the burden of persuasion and the burden of production. As it turns out, another Professor Cohen—Law Professor Neil Cohen—has already attempted to use the notion of sufficiency of information as an explanation of the gatecrasher and blue bus hypotheticals that accounts for the burden of persuasion and the burden of production. In doing so, Professor Neil Cohen has brilliantly expressed the idea of sufficiency of information on a level that is accessible to those not trained in formal mathematics. His ideas will be explored in the next section.

C. Professor Neil Cohen’s Confidence Principle

In his New York University Law Review article “Confidence in Probability,” Professor Neil Cohen advances the argument that the key to understanding why the defendants in the blue bus and gatecrashers hypotheticals should win is based on the idea of the sufficiency of information rather than probability. Professor Cohen concedes that “probability theory has a legitimate application to problems of legal proof,” but that “[s]cholarly debate [deriving from probability analysis] has failed to produce any satisfactory responses to [the blue bus and gatecrasher] examples.” According to Professor Cohen, the more information one has about an unknown fact, the more confident one can be that the probability assigned to that unknown fact is valid.

To demonstrate his point, Professor Cohen relies on the following hypothetical:

Suppose that you are asked to determine whether any of three coins is fair based on the following data: Coin A turned up heads 26,000 times in 50,000 flips; coin B turned up heads 27 times in 50 flips; and coin C turned up heads 51 times in 100 flips. Your best guesses, or point estimates, of the probabilities of heads for coins A, B, and C are 0.52, 0.54, and 0.51, respectively. From the point estimates, it might appear

---

155 See supra text accompanying note 137.
156 Cohen, supra note 14.
157 See id.
158 Id. at 392.
159 Id. at 395.
that none of the three coins is fair. However, the point estimates are not necessarily the same as true probabilities.\footnote{\textit{Id.} at 400 (footnotes omitted).}

In each of these three examples, the relevant data suggests that the coin might be “unfair” in the sense that the probability of heads and tails on each flip is not 50%. Further, each set of data suggests varying degrees of “unfairness” for each coin. However, the confidence one can have in the conclusion that each coin is unfair is different for each coin because the amount of information available for each coin is different. For coin A, which has been flipped 50,000 times, we can be relatively confident that it is, in fact, unfair. For coin B, however, which has only been flipped 50 times, we are less certain that it is unfair.

Professor Cohen’s point can also be made within the context of sporting events. Pretend that you are given the task of determining the likelihood that the Kansas City Royals won their baseball game on the previous night. Unfortunately, as it turns out, you do not have access to the morning paper (or any other such information) that would tell you the outcome of the previous night’s game. However, you do have access to accurate statistical information, updated on the morning of the game. For the first variation of this hypothetical, pretend that the statistical information reveals that the Royals had played 100 games, winning 60 and losing 40. If this were the only information you had regarding the game, you would have to peg the Royals’ chances of winning the previous night’s game at 60%. Naturally, though, more information, if available, would be preferred. Who were the Royals playing? What was the opponent’s record? Who was scheduled to pitch for the Royals? What was his record? What was his record against the opponent? Who was scheduled to pitch for the opponent? What was his record against the Royals? Was the game to be played in Kansas City or on the road? What was the Royals’ respective record at home or on the road? What was the opponent’s record at home or on the road? Was the Royals’ roster different for this game (perhaps because of injury or newly called-up players) than it had been for the majority of their previous 100 games? And so on and so on. Assuming that you
borrowed a computer used by Las Vegas sportsbooks in setting betting odds, you could eventually arrive at an assessment using all of the relevant data and variables. Ultimately, the assessment of a Royals victory on the previous night, based on the more extensive data, might end up being exactly 60%, which was the figure arrived at by considering only the Royals’ record to that point in the season. However, you would have more confidence in the 60% probability assessment if you factored in all of the specific relevant data that could be used in predicting the outcome. The process of considering all of the specific factors is more reliable than the process that only considered the Royals’ record.

According to Professor Neil Cohen, the reason that the defendants win in the blue bus and gatecrasher hypotheticals is the same reason we prefer the probability assessment of the baseball game that was based on the more specific information regarding location, opponent, pitchers, lineup, etc. Simply put, when the available evidence on a disputed question of fact is too generic, there can be very little confidence in any probability conclusion drawn from that evidence. The issue is not the probative value of the evidence; instead, the issue is that the relevant data is so scarce that our confidence in any probability assessment from that data is too low as to be acceptable:

In the hypothetical situations posited by Cohen and Tribe, the amount of information known about the incident in question is only a small fraction of all of the information that could be known about it. In other words, the estimated probability of the defendant’s liability is based upon only a small amount of data.

Because of this lack of information, Professor Neil Cohen asserts, the plaintiff has failed to show that there can be a “certain level of

---

161 The non-gambler might be amazed at the variety of variables considered by the sportsbooks in setting betting odds on sporting events. In order for the sportsbook to protect the advantage given to them by the standard 10% “juice,” it is imperative that its information be at least as accurate as anything available to the betting public. See Les Carpenter, *Betting Revolution Sweeps Vegas’ Gaming Industry*, YAHOO! SPORTS (Jan. 31, 2011, 5:01 PM), http://sports.yahoo.com/news/betting-revolution-sweeps-vegas-gaming-220100565--nfl.html (explaining how computers calculate betting odds).

162 Cohen, supra note 14, at 397-98 (footnote omitted).
confidence that the true probability, based on all possible evidence, exceeds [some] threshold.\textsuperscript{163}

The concept articulated by Professor Cohen is also evident in the ubiquitous public opinion poll, which is always accompanied by a margin of error. The margin of error simply represents the confidence that can be placed in those poll numbers.\textsuperscript{164} Obviously, if more people are polled, and if efforts are made to ensure that the poll is truly random, the margin of error shrinks.\textsuperscript{165} Using this analogy, Professor Cohen’s argument is that the margin of error in the gatecrasher and blue bus hypotheticals is so high—because the evidence is so generic—that insufficient confidence exists in the conclusion that the material fact necessary to the defendant’s liability actually occurred in the real world.

D. Confidence: Burden of Production Only or Burden of Persuasion Also?

Professor Neil Cohen’s confidence concept explains the blue bus and gatecrasher hypotheticals in a way that a probability analysis cannot. Even once the confidence principle is identified as the resolution to the paradox of the blue bus and gatecrasher hypotheticals, however, it is necessary to place the confidence principle within the framework of the burden of production and burden of persuasion. This purely legal inquiry involving the respective authority between judge and jury at the trial court level has been mostly ignored in the mathematical-driven discussion of the blue bus and gatecrasher inquiries. Professor Cohen offers a conclusion as to the appropriate location of his confidence principle, but his conclusions on this issue are less resistant to critique than his work in identifying the principle.

As used in this Article, the terms burden of production and burden of persuasion refer to the division of authority in the litigation context. The burden of production implicates the analysis a judge must perform in determining whether a case

\textsuperscript{163} Id. at 399.

\textsuperscript{164} See Barry Cushman, Mr. Dooley and Mr. Gallop: Public Opinion and Constitutional Change in the 1930s, 50 BUFF. L. REV. 7, 79-91 (2002) (discussing public opinions polls and margin of error in the context of the infamous Literary Digest poll pertaining to the 1936 presidential election).

\textsuperscript{165} See id.
should proceed to trial. Alternatively, the burden of persuasion implicates the analysis a jury conducts in determining whether to award judgment for the plaintiff. Under this dichotomy, Professor Cohen's confidence principle must be part of the judge's analysis of the burden of production. The "paradox" of the blue bus and gatecrasher hypotheticals involves a judge's entry of judgment for the defendant before the case ever reaches the jury. Professor Cohen's confidence principle only resolves this paradox if it is part of the judge's analysis of the burden of production.

Professor Cohen, however, identifies the confidence principle as being part of the burden of persuasion:

Not only must factfinders determine that their best estimate of the probability in question exceeds the threshold level—0.5 for the preponderance of the evidence standard—based on the evidence presented, but they also must have a certain level of confidence that the true probability, based on all possible evidence, exceeds that threshold. 166

Professor Cohen develops his theory that the confidence analysis is part of the burden of persuasion by distinguishing between "point estimates" of probability and "interval estimates":

[A] point estimate, which is an observer's best single estimate of a value such as the probability of liability, conveys much

---

166 Cohen, supra note 14, at 399; see also Neil B. Cohen, Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye, 73 CORNELL L. REV. 78, 86 (1987) ("Convincing the factfinder of such a probabilistic judgment requires more, I believe, than simply noting that the best guess of the probability exceeds 0.5; rather, I believe, the factfinder also takes into account its judgment as to how likely the best guess is to 'hold up."). Professor D.H. Kaye critiqued Professor Cohen's use of a confidence interval. See D.H. Kaye, Apples and Oranges: Confidence Coefficients and the Burden of Persuasion, 73 CORNELL L. REV. 54, 57 (1987) (advancing a complex mathematical critique of Professor Cohen's approach to the burden of persuasion and labeling Cohen's ideas as "incoherent"); see also Cohen, supra, at 82 (characterizing Kaye's critique as obscuring Cohen's central thesis that the burden of persuasion includes consideration of the amount of evidence available in the case). As will become evident later in this Article, I believe that Professor Cohen's confidence principle is useful to understanding the burden of production rather than the burden of persuasion. As such, it is unnecessary for me to venture into the debate between Professors Cohen and Kaye, which revolved entirely over how a confidence analysis could coexist with a probability analysis at the burden of persuasion stage. See Kaye, supra, at 77 ("I hope that my effort to spell out Cohen's theory and its implications will challenge Cohen and others . . . to provide a deeper theory of the burden of persuasion.").
less information than an interval estimate. An interval estimate goes beyond the “best guess” provided by the point estimate. It tells us how precise that guess is by describing a range of values within which one has a particular level of confidence that the true value lies. The point estimate, which does not indicate its precision, gives the user who is not aware of its nature a false sense of exactitude.167

Professor Cohen uses a “probability curve” to depict the jury’s confidence analysis. A probability curve requires a jury to do more than simply assign what it believes to be the most likely probability for the disputed question of material fact. Under a probability curve, the jury has to consider how much confidence it has in that probability assessment. In essence, a probability curve requires a jury to make a probability assessment of various different probabilities. A probability curve is depicted below:

![Figure I](image-url)

**Figure I**

In Figure I, the X-axis represents the spectrum of different probabilities that might be assigned to the disputed question of material fact. The Y-axis represents the likelihood that the probability assessment is the “true” probability of the existence of the disputed question of material fact. Thus, in this chart, the jury’s “best guess” from the available evidence is that there exists a 60% probability that the disputed question of material fact occurred. But, because the jury believes that there is additional information that might be helpful in assessing the probability of

---

167 Cohen, *supra* note 14, at 400 (footnote omitted).
the disputed event, and that this additional information might change its best-guess estimate of probability, it acknowledges that a different probability assessment is the “true” probability assessment. Thus, other probabilities are accounted for, reflected by the curve of the chart.

In situations in which the jury feels as if it has very little of the total amount of relevant information on the disputed question of material fact, the curve assigned by the jury will be flat. This is analogous to the Royals hypothetical, in which all that is known is the Royals’ record to that point in the season. The best guess, based on limited evidence, that the Royals won on the previous evening is 60%, but there can be little confidence in this conclusion because of the lack of information on which this estimate is based. This dearth of information requires an acknowledgement that the “true” probability of a Royals victory is very possibly something other than 60%. With more information, the best point estimate of the probability of a Royals victory might very well change. A flattening of the parabolic curve, as depicted below, reflects this relative lack of information:

![Figure J](image)

Figure J

Figure J, with its flat curve, is similar to the blue bus and gatecrasher hypotheticals, in which the point estimate of probability was based on a fraction of the information that would be relevant to establishing a probability assessment. When, however, more information is available (such as when the particulars regarding the Royals game are known—pitching
matchup, home or away, etc.), the likelihood that additional information would change the best-guess point estimate is reduced. A steepening of the curve, as reflected below, illustrates this situation.

![Figure K](image)

After determining the “pitch” or steepness of the curve (which is an exercise in determining the amount of relevant information produced and the likelihood that additional information would change the probability assessment), it is then necessary, according to Professor Cohen, to establish a permissible “confidence interval” or “margin of error.”168 “The difficult, but critical, question presented by the new model is ascertaining the level of confidence that the legal system should require in constructing the interval estimates used in the preponderance of the evidence test.”169 The critical question is this: What margin of error will the legal system tolerate that the probability assessment given by the jury would shift from a victory for the plaintiff (meaning, under the preponderance of the evidence standard, a >50% point estimate of probability) to a defeat for the plaintiff (a 50% or less point estimate of probability)? Once the acceptable risk of error is determined, the task of applying the confidence principle within the burden of persuasion is simply an exercise in determining

---

168 See generally id.
169 Id. at 410.
whether the remaining curve falls entirely above the 50% point associated with the burden of persuasion: “[W]e can be confident that the probability of the fact at issue is greater than 0.5 only if the entire interval estimate exceeds that value.”170

Thus, if a 95% confidence rate were required (meaning the acceptance of a 5% risk that the jury’s probability conclusion would change from plaintiff to defendant if more information became available), the confidence curve would look like this:

![Figure L](image)

Here, 95% of the curve is included and the 5% margin of error (colored gray in Figure L) is eliminated from the left tail of the curve.

A legal system that is more tolerant of the risk that the jury’s probability assessment would change with additional information will allow a higher margin of error. This might mean an increase in the margin of error to 25%, with a corresponding decrease of the confidence interval to 75%. In this situation, more of the portion of the curve is eliminated from consideration, as shown below:

---

170 *Id.* at 404.
Figure M

Under Cohen’s approach, then, the jury must conclude in Figure L that the burden of persuasion has not been met because of an unacceptable risk that additional information would change the jury’s best guess point estimate to the left side of the pivotal 50% mark. However, in Figure M, the jury should award judgment for the plaintiff because, as additional information became available, 75% of the time the jury would still assign a point estimate of probability to the right side of the 50% mark.

To complete Cohen’s description of how the confidence principle would work within the burden of persuasion, notice how the acceptable margin of error analysis functions in conjunction with the determination in each case as to how much of the relevant information is available. This latter determination establishes the pitch of the curve. Even assuming a static margin of error, whether the acceptable portion of the curve falls on the right side of the 50% mark will depend upon the pitch of the curve. So, in Figure L, the rather flat curve represents a case in which the factfinder believes that very little of the total probative evidence is available from which to make a probability assessment. This is analogous to the blue bus, gatecrasher, or the Royals hypothetical in which only the Royals’ win/loss record is known. But, if a factfinder believes that most of the total probative evidence is available, such as the Royals’ hypothetical in which a multitude of statistics are available (opponent, pitching matchups,
home/away, etc.), the pitch of the curve steepens. This steepening of the pitch can change whether the acceptable portion of the curve (defined by the margin of error) falls to the right of the 50% mark. Accordingly, in Figure N below, the same point estimate has been assigned from the available evidence as in Figure L (60%), and the same margin of error has been utilized as in Figure L (5% margin of error), but now the acceptable portion of the curve falls to the right side of the relevant 50% point because of the factfinder’s belief that most of the relevant information on the disputed question is available:

![Figure N](image)

Of course, in explaining the confidence principle as part of the burden of persuasion, Professor Cohen is not literally advocating for juries to engage in the “technical” discussion described above. Professor Cohen is not suggesting that juries do, or should, engage in a determination of the acceptable statistical “margin of error” analysis or that juries should formally chart the pitch of the confidence interval curve. Rather, Professor Cohen’s model serves a “heuristic . . . function” in that the model carefully describes the general manner in which a jury would consider the confidence question within the burden of persuasion. This is a significant and brilliant achievement.173

---

171 Id. at 417.
172 Id.
173 Other scholars have also concluded that a jury’s analysis under the burden of persuasion should include something akin to a “confidence” analysis. See Richard W.
There are, however, reasons to doubt that the confidence principle Professor Cohen describes should actually be part of the burden of persuasion analysis. Recall that the fundamental paradox of the blue bus and gatecrasher hypotheticals is that a judge would resolve the case by entering judgment for the defendant before the case ever reached the jury. Accordingly, the confidence principle identified by Professor Cohen must be part of the judge’s burden of production analysis, because it is this analysis that permits a judge to enter judgment before the case reaches the jury.

Professor Cohen’s inclusion of the confidence principle as part of the jury’s burden of persuasion analysis is not necessarily inconsistent with the notion that the confidence principle is also part of the burden of production. Recall that a probability analysis is currently part of both the judge’s burden of production analysis and the jury’s burden of persuasion analysis. Under the reasonable jury measure of summary judgment, the judge roughly replicates the probability analysis expected of the jury, but does so in a deferential way, requiring the judge to determine what range of probabilities would be “reasonable” from the record evidence. A similar relationship between judge and jury with regard to the confidence principle is theoretically possible. This would require the jury, at trial, to be instructed and informed regarding the confidence principle it is expected to perform in a similar fashion to how the jury is currently instructed regarding the probability principle. At the burden of production stage, a judge would

Wright, Liability for Possible Wrongs: Causation, Statistical Probability, and the Burden of Proof, 41 L.O.Y. L.A. L. REV. 1295, 1298 (2008) (concluding that the burden of persuasion in a civil case “requires the formation of a minimal degree of belief, based on evidence specific to the particular occasion, in the actual existence of the disputed fact in the particular situation”). And, regardless of whether, normatively, a jury should be expected to engage in a confidence analysis, there is empirical evidence suggesting that juries do, currently, engage in a confidence analysis as part of the burden of persuasion. See Gary L. Wells, Naked Statistical Evidence of Liability: Is Subjective Probability Enough?, 62 J. PERSONALITY & SOC. PSYCHOL. 739, 748 (1992) (describing results of a study suggesting that juries are uncomfortable basing decisions on pure statistical evidence); Dorothy K. Kagehiro & W. Clark Stanton, Legal vs. Quantified Definitions of Standards of Proof, 9 LAW & HUM. BEHAV. 159, 164-69 (1985) (describing the results of an empirical study demonstrating a divergence in results when sample groups are instructed under the preponderance of the evidence standard and when sample grounds are instructed under a quantified 51% probability standard).
determine whether a “reasonable” jury could determine that sufficient information exists from which to satisfy the confidence inquiry.

There are good reasons, however, to be skeptical that a jury is capable of performing the confidence analysis. A probability analysis simply requires a jury to draw upon experience and observation of witnesses within a courtroom, in order to assign a probability to a disputed question of material fact. Conversely, as detailed above, a confidence analysis requires a conclusion on two additional questions: The acceptable margin of error and the parabolic pitch.

The acceptable margin of error question requires a determination as to the legal system’s tolerance of the possibility that further evidence might change the probability assessment of a material fact. This is a policy or legal question that is outside the depth of the jury.174 The points made by Professors Zuckerman, Nesson, and Lempert, discussed above, address this policy issue of the acceptable margin of error. Their insights are not mutually exclusive; each of the policy considerations discussed by these Professors might influence the acceptable margin of error. Thus, the acceptable margin of error might include Professor Lempert’s concern that requiring more specificity will ultimately result in more specific evidence being produced.175 Additionally, the acceptable margin of error might also include a consideration of Professor Nesson’s concern that decisions based on scant evidence might erode public confidence in the system.176

Moreover, the acceptable margin of error might change depending on the type of case at bar and the substantive law that applies. Legislatures at both the federal and state levels frequently enact presumptions regarding the effect that can be given certain evidence.177 Usually, this works in favor of establishing a material fact, but if legislators can require that a

174 See Koehler & Shaviro, supra note 97, at 251-52 (stating that the “fact-finder’s confidence about the probability that particular facts are true” is a “policy concern[]” that is different than the question of “verdict accuracy”).
175 See supra note 150 and accompanying text.
176 See supra notes 146-48 and accompanying text.
certain fact be presumed from circumstantial evidence, it follows that they could do the same thing in reverse, which is to provide that certain circumstantial evidence is insufficient to establish a material fact and, thus, not in satisfaction of a party’s burden of production.

The objective here is not to identify all of the policy concerns that might influence the acceptable margin of error. The point is that a confidence analysis requires a margin of error determination, and the resolution of this issue involves policy matters much better suited to the judge rather than the jury.\footnote{Professor Cohen himself seems to implicitly recognize the difficulty associated with selecting an appropriate margin of error. After engaging in a very technical (and typically thorough and insightful) analysis as to the level of confidence that should be required, Professor Cohen concludes: “In sum, choosing the appropriate confidence level for the preponderance standard is not simple.” Cohen, supra note 14, at 417.}

The other question to be resolved as part of the confidence analysis requires a determination of how much of the available evidence has been presented and the likelihood that more information might change the best-guess point estimate of probability. This is the parabolic “pitch” inquiry discussed above. On this question, the jury, comparatively speaking, is in a better position than they were on the margin of error question. The pitch inquiry requires a determination as to the likelihood that additional information would change the probability conclusion; this inquiry is somewhat comparable to the probability analysis expected of the jury. Both require consideration of the probative value of evidence. Of course, a trial judge who has experienced hundreds or thousands of trials will probably be better aware of the types of evidence that might be available on disputed questions of fact. However, a judge, who has experienced hundreds or thousands of trials, is arguably in the best position to evaluate the probative value of the evidence actually adduced at trial to determine a point estimate of probability; yet, that decision remains within the province of the jury.

Even assuming the jury could handle the “pitch” question to determine how much of the potentially available evidence has been adduced at trial—and how likely it is that the additional information would change the point estimate of probability—the confidence analysis requires both the “pitch” analysis and the
margin of error analysis. As previously explained, the margin of error analysis appears to be a policy question beyond the capabilities of a jury. Further, there does not appear to be a feasible method by which the judge could first conduct this policy analysis and then communicate this standard to the jury. No one would seriously advocate the assignment of a mathematical margin of error expressed in statistical terms, and trying to articulate to the jury the appropriate margin of error in a verbal formulation seems impracticable.

In addition to the above points, the term “preponderance of the evidence,” which is the standard formulation given to a jury regarding the applicable burden of persuasion in a civil jury trial, supports the notion that probability, rather than confidence, is the only analysis expected from the jury. The term “preponderance” is defined as “superiority in weight . . . power, influence, importance, or strength.” Thus, by definition, the term “preponderance” suggests a weighing or comparing between two things. In the context of a civil trial, the thing to be weighed is “the evidence”—and theories from that evidence—that are submitted by the parties. Incorporating the notion of confidence into this verbal articulation of the jury’s burden of persuasion is difficult, to say the least.180

---

179 Webster’s Third New International Dictionary 1791 (1971).

180 See Kevin M. Clermont, Standards of Decision in Law: Psychological and Logical Bases for the Standard of Proof, Here and Abroad 122-23 (2013). (“[Cohen’s] approach is extraordinarily complicated . . . . One cannot specify the required level of confidence simply, no less picture it; human factfinders could never apply the approach, or even use it as a helpful image.”). This conclusion does not necessarily apply to either the “clear and convincing evidence” or “beyond a reasonable doubt” standard, both of which seem more susceptible to the notion that the jury is to apply a confidence analysis in addition to a probability analysis. First, the verbal expression of these standards, as opposed to the preponderance of the evidence standard, is more consistent with the notion that—to use Professor Cohen’s analogy—the jury is free to decide that there simply is not sufficient evidence from which to make a “bet” as to the defendant’s guilt or liability. See Kevin M. Clermont, Standards of Proof Revisited, 33 Va. L. Rev. 469, 477 n.15 (2009) [hereinafter Standards] (“[T]he culturally fraught criminal standard of proof and its wordy formulation could convey, on an emotional level, the message to the fact-finder’s unconscious that there is this need for personal conviction based on confidence in guilt, plus a very high probability of guilt.”); see also Stephen J. Fortunato, Jr., Instructing on Reasonable Doubt After Victor v. Nebraska: A Trial Judge’s Certain Thoughts on Certainty, 41 Vill. L. Rev. 365 (1996) (engaging in an extended analysis of the beyond a reasonable doubt standard, including a discussion of the confidence principle); cf. Cohen, supra note 14, at 421.
An analogy made by Professor Cohen nicely captures the issue.181 According to Professor Cohen, the burden of persuasion for a plaintiff in a civil trial can be analogized to a bet.182 Under this analogy, the jury is in a position to place a bet on the material facts in a case according to their assessment of the probabilities.183 If the jury finds the plaintiff’s evidence more persuasive, it can bet on the plaintiff.184 Similarly, if the jury finds the defendant’s evidence more persuasive, it can bet on the defendant.185 According to Professor Cohen, the jury has a third option, which is not to place a bet at all: “Under my view of legal factfinding, the party who has the burden of persuasion has the burden of convincing the factfinder to make the bet on that party. If the factfinder instead would bet on the other party, or choose not to bet, the burden has not been met.”186

Put simply, I believe that Professor Cohen has erred in concluding that the jury can refuse to bet based on a lack of confidence resulting from the dearth of information presented at trial.187 The better view is that a judge, as part of the burden of production analysis, should decide as a matter of law whether there is sufficient record evidence to allow the case to reach the

(suggesting that the clear and convincing evidence and preponderance of the evidence standards might differ in the level of confidence required). Second, when either of these two standards applies, the jury will usually be in a better position to understand the policy considerations that go into determining (or interpreting the judge’s instructions so as to apply) the appropriate level of confidence. See Charles R. Nesson, Reasonable Doubt and Permissive Inferences: The Value of Complexity, 92 HARV. L. REV. 1187, 1225 (1979) (“[A]ny conceptualization of reasonable doubt in probabilistic form is inconsistent with the functional role the concept is designed to play.”). See generally Fortunato, supra, at 373 n.35 (listing some of the leading sources on the “spirited debate” regarding “probability theory and its applicability to criminal jury trials”).

181 See generally Cohen, supra note 166, at 89-90.
182 See id.
183 See id.
184 See id.
185 See id.
186 See id. at 90.
187 Here again, it is helpful to differentiate between whether a jury should be allowed to engage in a confidence analysis (a normative question) and whether juries do, in fact, engage in a confidence analysis. My point here is only with the normative question. There is some interesting empirical evidence suggesting that, on the descriptive point, juries do sometimes engage in a confidence analysis. See supra note 173; see also Standards, supra note 180, at 482 (“Regardless of the desirability of looking to probability, it could be that many fact-finders in actuality look to confidence when called upon to apply a standard of proof.”).
jury. Then the judge decides whether a wager will be placed. The jury decides how to place that bet, and they are not free to decide not to bet at all.

If I am correct that the confidence analysis is part of the judge's analysis (burden of production) and not part of the jury's analysis (burden of persuasion), altering Professor Cohen's depiction of the confidence analysis becomes necessary. Professor Cohen's model of the confidence analysis is based on the presumption that the same entity—the jury—would be performing both the probability analysis (assigning a point estimate of probability) and the confidence analysis. However, this model does not work as well in describing the judge's analysis of the confidence question within the burden of production. Although a judge does consider probability as part of his analysis of the burden of production, a judge's probability analysis does not require him to assign a point estimate, but rather to determine a range of probabilities within which a point estimate would be "reasonable." The judge's analysis of probability at the burden of production stage was depicted earlier in this Article and is reproduced in Figure O. Recall that in Figure O, the gray area represents the range of point estimates that are a reasonable conclusion from the given estimate.
Thus, depicting the judge’s analysis of confidence at the burden of production stage (and assuming that the jury does not engage in a confidence analysis) necessitates incorporating the confidence analysis into the preexisting chart that depicts the judge’s deferential analysis of the probability question. Although Professor Cohen’s technical statistical depiction accurately captures the confidence analysis, this level of technicality is perhaps not necessary to convey the general concept that a judge

---

188 It might be tempting to conclude, from the depictions used to this point, that Professor Cohen’s confidence analysis is really the same as Professor McNaughton’s depiction of the reasonable jury inquiry. Both are charted similarly; both have a range of point estimates of probability that spread out over an X-axis. In reality, however, a confidence interval such as the one used by Professor Cohen and the reasonable jury range depicted by Professor McNaughton are measuring entirely different things: probability and confidence. This is perhaps best illustrated by considering how the blue bus hypothetical would be depicted under each approach. Under Professor McNaughton’s depiction, the range of reasonable probability estimates would be nonexistent. Under the blue bus hypothetical the only available evidence suggests that there is an 80% probability that the defendant is liable. Assuming that a jury is only to consider probability, and not confidence, a jury would be unreasonable in pegging the probability of the defendant’s liability at anything other than 80%. So, the range of the reasonable point estimates of probability that are possible from the record evidence reduces to exactly the 80% mark on the chart. Under Professor Cohen’s confidence interval, however, there can be very little confidence that the 80% point estimate of probability is the “true” estimate of probability. Because so little of the available information as to the identity of the bus is known, the confidence interval “flattens” such that it extends over a large portion of the X axis. Thus, in a case such as the blue bus, the range of values represented by Professor McNaughton’s depiction of the judge’s probability analysis shrinks to exactly 80%. On the one hand, Professor Cohen’s confidence interval flattens and spreads to reflect that, although 80% is the best point estimate of liability from the existing evidence, there is a strong possibility that that point estimate of liability would change as more information became available.
must be assured that there is enough evidence on the disputed question of fact to allow a minimum level of confidence in the jury’s probability assessment.

To chart this inquiry, the previous one-dimensional chart of Professor McNaughton, simply needs to be expanded into a two-dimensional inquiry. The X-axis will continue to depict whether the reasonable jury measure of the burden of production has been met. As before, the relevant consideration is whether the gray area (representing the range of reasonable probability) extends on both sides of the 50% mark (representing the usual burden of persuasion at trial). If so, the probability inquiry within the burden of production has been satisfied because, at trial, a reasonable jury could resolve the probability question in favor of either the plaintiff or the defendant. To additionally depict the judge’s evidentiary specificity analysis, a Y-axis must be included. The Y-axis measures the specificity of the evidence adduced, with an imaginary line representing the minimum amount of factual specificity required for the plaintiff to survive a summary judgment motion:

Figure P

The judge’s dual task in determining whether the parties have met their respective burdens of production is made clear by Figure P. First, the judge must assign a range of probability values on the X-axis by conducting the reasonable jury inquiry. The judge does this by considering the range of probability

---

189 The risk that the confidence requirement will not be met falls on the party with the burden of persuasion at trial, which will usually be the plaintiff. See supra note 70 and accompanying text.
assessments that are reasonable from the record evidence. As long as the range extends over the relevant burden of persuasion, both parties have met their obligation under the reasonable jury inquiry. The judge must also consider whether the available evidence is sufficient to rise to the level required under the confidence analysis. This is charted on the Y-axis. The risk that this confidence standard has not been met falls exclusively on the party with the burden of persuasion, which will usually be the plaintiff. Thus, below is depicted a case in which the trial court judge believes that the scant summary judgment record would support a reasonable probability conclusion for either party but that the record is simply too sparse to allow the plaintiff to proceed to trial. As such, the vertical measure of confidence (measured on the Y-axis) fails to rise to the level of confidence the judge perceives as being required—pursuant to a legal and policy inquiry—for that particular case.

In Figure Q, the plaintiff has satisfied the reasonable jury inquiry, but has not satisfied the confidence requirement. Accordingly, the judge must enter judgment for the defendant.

If the plaintiff has direct evidence on each material fact, the plaintiff will satisfy this element of the burden of production. See supra notes 94-95 and accompanying text.
CONCLUSION: THE CONFIDENCE CONCEPT AND THE REASONABLE JURY STANDARD

The lesson from the previous Section is that Professor Neil Cohen was correct to identify the confidence principle as the explanation for the “paradox” of the blue bus and gatecrasher hypotheticals, but that he erred in assuming that a confidence analysis was part of the jury’s analysis of the burden of persuasion. Because a confidence analysis requires a legal or policy decision as to whether there is an adequate quantum of evidence in a particular dispute, the judge must make this determination as part of her analysis of the burden of production.191

Completely understanding the paradox of the blue bus and gatecrasher hypotheticals, then, requires one to draw two distinctions. First, the distinction between probability and confidence. This distinction might be a difficult one for many lawyers to make, considering the profession’s general resistance to mathematical concepts. Second, the distinction between judge and jury. This division of authority within the trial court is familiar enough to most lawyers. The reasonable jury standard, however, misrepresents the division of authority at the trial court level between judge and jury, with regard to the confidence principle. Further, the process of making the identification of the confidence principle, which is challenging enough, becomes even more difficult.

The reasonable jury standard presupposes that a judge’s analysis of the burden of production essentially replicates the jury’s analysis of the burden of persuasion. But, as discussed above, it seems unlikely that the legal system can expect a jury to ever engage in a confidence analysis. Because the confidence analysis requires a legal or policy determination, the confidence principle should be assigned to a trial court judge only and not to the jury.192 The reasonable jury standard, however, by defining

191 Indeed, the paradox of the blue bus and gatecrasher hypotheticals is not resolved until the judge’s entry of judgment for defendant, before a jury trial, can be explained. The judge’s analysis under the burden of production is thus the only route by which the paradox of the blue bus and gatecrasher hypotheticals can be solved.

192 Even if Professor Cohen is correct in assuming that the jury is also to engage in a confidence analysis, the reasonable jury interpretation of Rule 56 clouds the
the judge’s burden of production analysis with reference to the jury’s burden of persuasion analysis, makes it nearly impossible to identify this already slippery concept as part of the judge’s pre-trial inquiry. Instead, the reasonable jury standard precipitates the notion that the judge and jury are engaged in the same inquiry. Equating the judge’s and jury’s inquiries leads to statements like the following:

The distinction between a burden of production and the burden of persuasion is a matter of degree rather than kind, and the difference between the two burdens should not be overstated. In order to determine whether a burden of production is satisfied, one must evaluate the persuasive force of the evidence that has been adduced on the relevant issue, and a litigant has met his burden of production only if the evidence introduced is sufficiently persuasive to create a reasonable doubt, or to meet some lower standard.193

This quote is accurate with regard to how probability is decided within the litigation context: The jury is the primary decision-maker, while the judge acts as a sort of gatekeeper to resolve truly obvious cases. The reasonable jury standard accurately captures this relationship between judge and jury with identification of the confidence principle within the burden of production. The reasonable jury language suggests that a judge is simply to mimic the analysis of the jury, albeit in a deferential manner. It is not clear that the same deference that the judge owes to the jury with regard to probability is also warranted under Professor Cohen’s vision in which both the judge and the jury perform a confidence analysis. Even more fundamental, however, is the close association between the jury and a probability analysis. The burden of persuasion, in the mind of most practicing attorneys and judges, is about probability. Therefore, it is noteworthy when someone suggests, as Professor Cohen has done, that the jury’s responsibility under the burden of persuasion extends beyond the assignment of point estimate probabilities. Even if Professor Cohen is correct in suggesting that a jury’s role includes both probability and confidence, defining the burden of production by reference to what a “reasonable jury” could conclude does little to demarcate the separate and distinct analysis required by a confidence analysis. As a practical matter, because most people associate the function of a jury with probability, efforts to establish the confidence principle as a distinct concept within the burden of production (and, even, within the burden of persuasion under Professor Cohen’s view) would benefit from a rejection of both the “reasonable jury” language and all the baggage associated with that term.

regard to probability. The reasonable jury standard utterly fails, however, to convey that a confidence inquiry is an additional route by which a lawsuit might be resolved. The reasonable jury standard conceals the confidence principle.\textsuperscript{194} 

\textsuperscript{194} The astute reader will note that although the reasonable jury standard makes identification of the confidence principle more difficult, I have not called for an elimination of the reasonable jury standard (which, as noted at the outset of this Article, is a judicial gloss on the language of Rule 56). Before this argument can be advanced, it is necessary to determine whether the Federal Rules even sanction the use of the confidence principle. This important topic has not been addressed herein, but should be evaluated in the future. Obviously, if the Federal Rules do not sanction the type of confidence inquiry that resolves the paradox of the blue bus and gatecrasher hypotheticals, it is less problematic that the reasonable jury standard makes the confidence analysis more difficult to discern.