Faculty Publications

2001

Naturalized Epistemology and the Critique of Evidence Theory

Dale A. Nance
Case Western University School of Law, dale.nance@case.edu

NATURALIZED EPISTEMOLOGY AND THE CRITIQUE OF EVIDENCE THEORY

*Dale A. Nance**

INTRODUCTION

IN their thought-provoking contribution to this conference,[1] Professors Ronald J. Allen and Brian Leiter address the state of evidence theory and purport to illustrate how it sometimes neglects developments in modern epistemology. They advance the thesis that "naturalized epistemology," and "in particular that branch of naturalized epistemology known as social epistemology," "provides the most appropriate theoretical framework for the study of evidence."[2] Actually, Allen and Leiter maintain that the approach of

---

* Professor of Law, Chicago-Kent College of Law, Illinois Institute of Technology.
[1] Ronald J. Allen & Brian Leiter, Naturalized Epistemology and the Law of Evidence, 87 Va. L. Rev. 1491 (2001).
[2] Id. at 1493.

social epistemology explains what most evidence law scholars in fact do, "regardless of their explicit philosophical commitments,"[3] a claim which I take to mean that most evidence scholars follow the general approach of modern naturalized epistemology, whether or not they are aware of that fact. If so, then such evidence scholars are admirably, if fortuitously, in tune with developments in modern philosophy. But problems arise, say Allen and Leiter, when we come to certain theoretically oriented evidence scholars who fail to recognize or to internalize appropriately the insights of social epistemology.

Allen and Leiter quickly pass over the work of those evidence theorists most concerned with "postmodern" conceptions of truth and knowledge,[4] and so will I. Their principal targets, which they summarily characterize as those in search of an "algorithm" for decisions at trial, are expected utility theory, Bayesian decision theory, and microeconomics.[5] These theories seem to have the following elements in common: They are wide ranging in what they purport to explain or rationalize; they involve—at least in principle—quantification of probabilities and (sometimes) costs and benefits; and they are formal, in the sense that they attempt to model decisionmaking by simplifying the structure of the task at hand to a form that permits quantitative manipulation. The first of these features does not appear to be the culprit, for in the course of their paper, Allen and Leiter advance their own fairly comprehensive theory of evidence law—the "relative plausibility" theory— primarily a theory about burdens of proof but with claimed implications for the principles of admissibility as well.[6] The real matters of concern appear to be the explicit quantification of probabilities and costs and, perhaps especially, the simplification of reality necessary to utilize formal models.

According to Allen and Leiter, algorithmic approaches—or at least the ones they criticize—neglect two important insights of naturalized epistemology.[7] First, prescriptions generated by any useful theory of evidence should comply with a general constraint:

---

[3] Id.
[4] Id. at 1492 & n.1.
[5] Id. at 1492–93.
[6] Id. at 1527–49.
[7] Id. at 1498.

"Ought implies can." "In other words, normative epistemology, like normative ethics, cannot require of agents actions (mental or physical) that they cannot perform."[8] Second, naturalized episte-mology is *instrumental*: Cognitive procedures should be assessed in terms of how effective they are in producing knowledge. In the context of the design of rules governing trials, this means that such rules should require only those cognitive tasks that triers of fact, often lay jurors, are capable of performing, and that such rules should be assessed in terms of whether they produce true belief about the litigated facts,[9] recognizing of course that some rules have a different purpose.[10]

In what follows, I give Allen and Leiter a mixed review. I ap-plaud their focus on naturalized epistemology, but I question the claims that they argue follow from it. In some ways, my reaction is that they have not gone far enough in pressing its implications, and I attempt to suggest how further progress might be made along this path. On the whole, I conclude that the antipathy toward algo-rithms expressed by Allen and Leiter is misplaced.

Part I will review the relationship between naturalized episte-mology and what they refer to as "The Conceptual Foundations of Evidence."[11] I will argue that the implications of naturalized epis-temology are not as supportive of a paternalistic attitude toward lay juries as Allen and Leiter appear to believe. Part II will discuss burdens of persuasion and the competing perspectives on such burdens provided by expected utility theory, which Allen and Le-iter reject, and the relative plausibility theory, which they offer in its place. Here I will argue that Allen and Leiter's preferred theory is best understood as a tool that can be utilized under a decision criterion better informed by expected utility theory. Part III will address their criticisms of Bayesianism. I will reply that Bayesian thinking has a legitimate role to play as long as we do not expect too much of it. I will not consider separately what Allen and Leiter have to say about microeconomics as a tool for analysis of evi-dence, because this part of their paper is singularly directed at Judge Richard Posner's interesting, but I think largely unproduc-

---

[8] Id. at 1499.
[9] Id. at 1499.
[10] Id. at 1500.
[11] Id. at 1499–1503.

tive, foray into the field of evidence,[12] replicating many of the same issues addressed in connection with expected utility theory and Bayesianism.[13] Along the way, however, I will address certain matters relating to Posner's arguments that shed light on the general issues considered.

I. NATURALIZED EPISTEMOLOGY AND THE CONCEPTUAL FOUNDATIONS OF EVIDENCE LAW

Allen and Leiter argue that modern epistemology is heavily influenced by the idea that "philosophy should be continuous with a posteriori inquiry in the empirical sciences," that "philosophy cannot be an exclusively a priori discipline."[14] Nowhere is this more plausible than in doing philosophical work about activities or enterprises aimed, at least in large part, at the creation of knowledge, such as science or trials in law courts. Naturalizing epistemology entails giving attention to the empirical study of knowledge creation. Social epistemology, in particular, is "that branch of naturalized epistemology concerned ... with the social processes and practices that inculcate belief."[15] In its normative or regulative dimension, social epistemology analyzes and evaluates social practices in terms of the objective of inculcating knowledge rather than ignorance, true rather than false beliefs.[16]

This kind of normative social epistemology obviously resonates well with what evidence scholars have been doing for generations, analyzing rules of proof in terms of what Allen and Leiter, following philosopher Alvin Goldman, would call their "veritistic" tendency—their tendency to generate true beliefs by decisionmakers about disputed facts.[17] Moreover, it certainly encourages the evidence scholar to take seriously what empirical social science has to say that is relevant to the design of evidence rules. Allen and Le-

---

[12] Richard A. Posner, An Economic Approach to the Law of Evidence, 51 Stan. L. Rev. 1477 (1999).

[13] Economic analysis of evidence law in general, and Posner's arguments in particular, are considered at length in another paper presented at this conference. See Richard Lempert, The Economic Analysis of Evidence Law: Common Sense on Stilts, 87 Va. L. Rev. 1619 (2001).

[14] Allen & Leiter, supra note 1, at 1494 (footnote omitted).

[15] Id. at 1497 (footnote omitted).

[16] Id. at 1498.

[17] Id.

iter illustrate the benefits of doing so by discussing empirical work on the evaluation of witness demeanor,[18] on the ability of jurors to understand the random match probability in DNA evidence cases,[19] and on the usefulness and prejudicial nature of character evidence.[20] While there has long been recognition of the importance of empirical work to evidence law and scholarship, there is an understandable reluctance among lawyers, most of whom have no scientific training, to dig into the details of empirical studies. Nevertheless, in principle, there is no incompatibility here, and one can readily agree with Allen and Leiter's claim that "[f]or the great bulk of evidentiary scholars ... this article merely solidifies the ground beneath their feet."[21]

Significantly, Allen and Leiter maintain that endorsing the veritistic framework of naturalized epistemology does not itself require them to take sides in the debate between

> those who advocate the "jury control principle" (the idea "that the organizing principle of Evidence law [is] a fear that lay jurors [will] misuse certain types of evidence") and those who advocate "the best evidence principle" (the idea that "[t]he best evidence must be given of which the nature of the case permits").[22]

Both views, Allen and Leiter correctly point out, entail a commitment to the design of veritistic institutions for the trial of disputes, specifically evidence rules. Still, the way Allen and Leiter express the debate is less than maximally illuminating in this context; it compares apples with oranges by juxtaposing the "jury control principle" with the "best evidence principle." More illuminating here would be to contrast the "worst evidence principle"—the idea that evidence law seeks to prevent jury error by filtering out the really bad evidence that is likely to lead the jury astray but otherwise leaves the parties free to choose their proofs—with the "best evidence principle"—the idea that evidence law seeks, within the context of an adversarial trial, to assure that the best evidence

---

[18] Id. at 1539–42.
[19] Id. at 1542–45.
[20] Id. at 1546–49.
[21] Id. at 1493.
[22] Id. at 1501 (footnotes omitted, alterations in original).

reasonably available to the parties is ultimately presented for the jury's use.[23] Expressing the contrast in this way reveals an important difference between the two views: The worst evidence principle emphasizes *jury control,* while the best evidence principle emphasizes *advocate control,* which may be represented as follows:

worst evidence principle ⇓	versus	best evidence principle ⇓
jury control principle	versus	advocate control principle
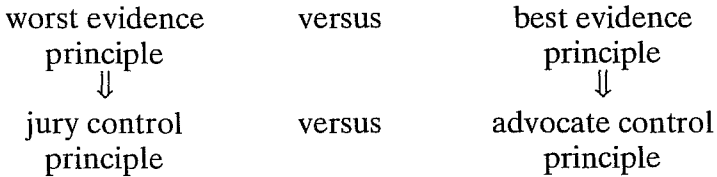
Why is this important? Despite their ostensible neutrality, Allen and Leiter assert that "[t]he jury control principle does, however, highlight an interesting feature of our evidentiary rules, namely, their epistemic paternalism."[24] This assertion warrants careful scrutiny. If their claim is that epistemic paternalism is significant *to the extent* that the jury control principle rightly interprets our evidentiary rules, then I have no quarrel with what is claimed, only with what is thereby omitted. If, however, their point is to assert the *dominance* of epistemic paternalism as the rationale for our evidentiary rules, then Allen and Leiter may indeed be taking sides.[25]

---

[23] Compare Dale A. Nance, The Best Evidence Principle, 73 Iowa L. Rev. 227 (1988) (making the case for the best evidence principle), with Edward J. Imwinkelried, The Worst Evidence Principle: The Best Hypothesis as to the Logical Structure of Evidence Law, 46 U. Miami L. Rev. 1069 (1992) (suggesting that the common law of evidence focused on preventing witness perjury). Note that the worst evidence principle expressed in the text above is somewhat different from the principle by the same name expressed by Professor Imwinkelried, who focuses only on the prevention and detection of perjury and does not explicitly rely upon the tendency to deceive jurors as the source of the concern to suppress perjury. For those unfamiliar with this debate, emphasis on the worst evidence principle, as articulated here, traces back to the work of James Bradley Thayer in the late nineteenth century, while emphasis on the best evidence principle traces back through the work of Edmund Morgan in this century to several treatise writers of the eighteenth and early nineteenth centuries. John Henry Wigmore, the ultimate veritistic evidence scholar, was ambivalent on the matter. Ample citations may be found in Nance, supra.

[24] Allen & Leiter, supra note 1, at 1502.

[25] Neither proponents of the jury control model nor proponents of the advocate control model maintain that their preferred perspective explains *all* evidence rules; everyone recognizes some scope for the opposing model. The differences are a matter of emphasis. Thus, it is not entirely clear from the passage quoted in the text or the rest of their paper whether Allen and Leiter do in fact take sides in this debate.

At the least, one would expect that they would also explore the contours of naturalized epistemology in the advocate control framework. The remainder of this Part addresses this matter.

As Allen and Leiter describe it, "[e]*pistemic* paternalism substitutes the rulemaker's judgment about what is *epistemically* best for the agents for their own judgment."[26] To the extent that a given rule can and should be interpreted in terms of advocate control, it would be misleading to speak in these terms. An advocate control explanation views the judge as agent for the jury, able to use courtroom powers rightly centralized in the judge (for other practical reasons) to protect juries from the epistemic consequences of third-parties' choices—namely, the choices of the advocates about what evidence to present and how to present it. In doing this, the judge need not assume any difference between his or her factual inferences and those of a jury, given the same information. The rules allowing (and enforcing) pretrial discovery of evidence known to an opponent are straightforward and important examples.[27]

It is true that the trial judge's judgment is always substituted for the judgment of the jury when it comes to the question of whether to *admit* evidence, but that is not necessarily the result of any substitution of judgment about what to infer about the merits of the case. Privilege rules are illustrative. We exclude evidence of privileged communications not because of any substitution of the judge's assessment of what is epistemically best for the jury—that is, best relative to their factfinding task—but rather because of ancillary policies.[28] More generally, the allocation of authority to the judge to exclude evidence is mostly a matter of administrative convenience and, of course, the superior legal knowledge of the judge. In those contexts governed by the advocate control principle, however, the judge's assessment about what evidence is practicably presentable is sometimes substituted for the jury's assessment of what is practicably presentable.[29] To the extent that the latter con-

---

[26] Id. at 1502 (quoting Brian Leiter, The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make for Good Philosophy of Evidence, 1997 BYU L. Rev. 803, 814–15).

[27] E.g., Fed. R. Civ. P. 26.

[28] Graham C. Lilly, An Introduction to the Law of Evidence § 9.1, at 437 (3d ed. 1996).

[29] For example, some substitution may be required to assess whether a hearsay declarant incarcerated in another state is practicably presentable by the prosecution;

sideration is involved, there is an element of paternalism in placing this authority in the judge rather than the jury. Its reason is epistemic in that it *relates to* evidence, but it is not purely epistemic—it involves a heavy dose of policy—and it is otherwise quite distinct from the kind of paternalism generated by a differential assessment of evidence admitted on the merits of the case.

The point can be made more precise with the following formal representation.[30] Any veritistic approach to admissibility presupposes that one can, always in principle and sometimes in practice, associate with each alternative evidence package that a jury might consider in a case a measure or ranking of the probable accuracy of the jury's verdict. Let $V(x)$ represent this association. Let E be the set of all evidence admitted or to be admitted, not including evidence E_o (for the "offered" item of evidence), the admissibility of which is in question. Under the worst evidence principle, and its associated jury control model, the evidence E_o should be excluded if the legal system believes[31] that

(Taint Criterion) $V(E) > V(E \text{ and } E_o)$.

In many applications, this criterion is paternalistic in the epistemic sense characterized by Allen and Leiter. If the legal system believes this inequality holds, it may be because officials within the legal system believe the jury will be unable to ignore or adjust for the truth-distorting effects of E_o, effects that by hypothesis must be better understood by those who would exclude E_o from the jury's

---

these are matters that the judge likely understands better than the jury. Cf. Barber v. Page, 390 U.S. 719 (1968) (holding that a witness is not "unavailable" unless the prosecution makes a good faith effort to obtain him for trial). It generally does not take much substitution of judgment, however, to appreciate that cross-examinable testimony is superior and reasonably presentable, as compared to testimony by a witness who would prefer not to be subject to cross-examination. See Nance, supra note 23, at 282.

[30] Presumably, this is a use of formalism to which even Allen and Leiter would not object.

[31] The expression "the legal system believes" and similar expressions are intended to prescind questions of allocation of exclusionary authority among trial courts, appellate courts, and legislatures. The belief in question may be that of the trial judge acting within appropriate discretion as to a particular case, or that of a rulemaking body acting with regard to a general class of cases into which E_o falls.

consideration.[32] In other applications however, the veritistic improvement attributable to excluding E_o when the Taint Criterion is satisfied can result from non-paternalistic factors, factors that would warrant exclusion even if those officials making the assessment under the criterion were also the triers of fact.[33]

Under the best evidence principle and its associated advocate control model, the evidence E_o should be excluded not only in the latter context—when excluding "tainted" evidence does not entail endorsing a seriously paternalistic attitude toward jurors[34]—but also when the law believes that there exists alternative evidence not within E—call it E_a—that the proponent could, and likely will, present if E_o is excluded and

(Inducement Criterion) $V(E \text{ and } E_a) > V(E \text{ and } E_o)$.[35]

In such cases, the law excludes the evidence E_o not because the jury will be unable to discount E_o appropriately but rather because it should not have to decide the case without the benefit of the better evidence, E_a. That is, under this theory, exclusion would be warranted even if it were also true that $V(E \text{ and } E_o) > V(E)$, in other words, even if E_o would not be excluded under the Taint Criterion. Conversely, if E_o is relevant, exclusion ordinarily would not

---

[32] For example, if E_o is hearsay, it might be thought that the jury would be unable to account for the epistemic hearsay dangers, such as the fact that the original declarant did not speak under oath subject to cross-examination (assuming that is the case), dangers known to the legal system and for which more knowledgeable individuals (for example, those with legal training) could make appropriate discounts.

[33] The conspicuous example is excluding evidence that is irrelevant or of such little probative value as to be a waste of time. See Fed. R. Evid. 402–03. Regardless of the level of epistemic competency of the trier of fact, there is something to be gained by being able to focus cognitive energy on evidence that is of significant probative value. The best evidence principle, at least as I have developed it, allows for taint-based exclusions of this kind. Nance, supra note 23, at 270–74.

[34] As I have elaborated it, the best evidence principle does not endorse the exclusion of evidence simply because of a belief that the jury will not understand the evidence or will use it in a flawed manner. For a discussion of these issues, see Nance, supra note 23, at 271.

[35] Here also, the probabilistic assessment of whether the proponent will introduce the alternative evidence may be one regarding the particular case or one regarding a class of cases of which the present case is an example.

be warranted if no such alternative evidence (E_a) could be identified.[36]

What is striking about the difference between these two veritistic admissibility criteria is that the Taint Criterion seeks to shield from the jury evidence that the jury is thought incapable of handling to good effect. By contrast, the Inducement Criterion excludes evidence only to improve the total evidentiary package by motivating a party to introduce evidence that would otherwise be omitted or that, if the alternative evidence is likely to be introduced by the opponent anyway, would not thereby be introduced in the most convenient way for the jury's use. The latter theory will rarely be paternalistic in any strong sense because it typically reflects the law's belief that any rational decisionmaker, whether lay jury or experienced trial judge, would benefit from the superior evidence presentation.[37] The difference between these two ideas is clearly reflected, but also underutilized, in the philosophical literature on epistemic paternalism upon which Allen and Leiter presumably draw.[38]

---

[36] For example, if E_a is otherwise admissible testimony concerning the contents of a document, exclusion of such testimony for failure to introduce the original document—here the E_o—only makes sense when it is practical for the proponent to introduce the original; otherwise, the decisionmaker benefits from having at least secondary evidence of the contents (E_a) the probative value of which must be duly discounted. Of course, this specification fits the pattern of practice actually encountered under the so-called "original document rule" much better than the Taint Criterion. See Fed. R. Evid. 1001–08.

[37] Whether a particular exclusionary rule should be understood, and potentially reformed, under a Taint Criterion model or an Inducement Criterion model is often a challenging question. See, e.g., Nance, supra note 23, at 274–94 (arguing that the Inducement Criterion model better explains many of the rules). The hearsay rule is particularly interesting and important in this regard. See id. at 281–83 (arguing that while hearsay rules have often been understood under a "jury distrust" model and possess some features reflecting that fact, they are actually better interpreted as an application of the "best evidence" principle); see also George F. James, The Role of Hearsay in a Rational Scheme of Evidence, 34 Ill. L. Rev. 788, 791–97 (1940) (arguing that the "best evidence" idea is the only *proper* ground for the exclusion of hearsay); Michael L. Seigel, Rationalizing Hearsay: A Proposal for a Best Evidence Hearsay Rule, 72 B.U. L. Rev. 893 (1992) (articulating steps to reform the hearsay rule in accordance with the best evidence principle).

[38] See Alvin I. Goldman, Epistemic Paternalism: Communication Control in Law and Society, 88 J. Phil. 113 (1991), *reprinted in* Liaisons: Philosophy Meets the Cognitive and Social Sciences 209 (1992). When addressing the exclusion of evidence from trials at law, Professor Goldman speaks only in terms corresponding to the taint criterion. Id. at 209–15. Later, when addressing Federal Trade Commission remedies

With these points in mind, we can better assess the significance of Allen and Leiter's warning that the epistemic paternalism of the jury control model, if it is to be done well, requires attention not only to the epistemic shortcomings of jurors but also to the epistemic shortcomings of the judges who must apply any rules designed to deal with the former.[39] This too can be granted without taking sides in the dispute over the relative merits of jury control and advocate control. Indeed, this same distinction, between what Allen and Leiter call "primary epistemic rules" and what they call "secondary epistemic rules,"[40] also applies in the context of advocate control explanations of evidence rules. Had Allen and Leiter explored this alternative path of analysis, they might have reached a conclusion more supportive of the advocate control view.

Most importantly, policymakers who write rules of evidence and judges who must apply them (both overwhelmingly lawyers) are by training quite familiar with what lawyers do in thinking about how to prepare and present a case. By contrast, lawyers are perhaps too professionalized to empathize effectively with the situation of lay jurors. Indeed, empirical studies conducted in recent decades have shown that many of the rather condescending assertions lawyers have been making regularly for decades, indeed for centuries, about how jurors react to evidence are false, or at least considerably off-target.[41] Put simply, we lawyers—including judges—know a

---

for deceptive advertisements, he recognizes that in corrective advertising "communication control sometimes takes a stronger form than the one considered thus far: not excluding messages, but mandating messages of a specified kind or content." Id. at 216. Of course, the two can be combined, as when the point of excluding a message is to encourage an alternative one, and this is the pattern often seen in the law of evidence.

[39] Allen & Leiter, supra note 1, at 1502.

[40] Id.

[41] One illustration is documented in a paper presented at this conference. Shari Seidman Diamond & Neal Vidmar, Jury Room Ruminations on Forbidden Topics, 87 Va. L. Rev. 1857, at 1875–95 (2001) (reporting the results of a study of actual jury deliberations indicating that jurors in tort cases who discuss the forbidden topic of insurance almost invariably discuss not the deep pockets of the insured defendant—so feared by lawyers as to have generated a codified response, Federal Rule 411—but rather the risk of double recovery by the insured plaintiff). There are many other such examples in the literature, including the accumulating empirical evidence that jurors are not overly impressed by hearsay evidence despite the persistent fears of lawyers that are offered to explain the hearsay rule. See, e.g., Peter Miene et al., Juror Decision Making and the Evaluation of Hearsay Evidence, 76 Minn. L. Rev. 683, 686–87 (1992); Richard F. Rakos & Stephan Landsman, Researching the Hearsay Rule:

lot more about what lawyers do and why they do it than we know about what juries do and why they do it. Thus, the factual assumptions used by judges and other policymakers in order to substitute their own judgments about inferences from evidence for those of juries are inherently less reliable than the experiential basis they use to substitute their judgments for those of advocates as to the presentation of evidence.

Such comparative institutional competency should be important to theorists who rightly emphasize the principle that "ought implies can."[42] It provides a strong reason to believe that rules designed to prevent or offset the accuracy-jeopardizing choices of advocates are more likely to yield veritistic fruit than are rules designed to prevent or offset jury error that might arise from lay jurors' epistemic incompetence. At least, this will be true until and unless particular suspicions about such jurors' epistemic deficiencies are confirmed by empirical research. Except to the extent that such deficiencies can be empirically confirmed, admissibility rules and other rules of trial procedure should be focused on empathetically assisting jurors to perform the difficult inferential task that they are assigned rather than on paternalistically suppressing information that lay people are thought incapable of using properly.

It is hard to discern from what Allen and Leiter have written whether correcting for these sins of omission (failure to explore the comparative ability of jurists to effectuate jury control and advocate control) and commission (overemphasis of epistemic paternalism) would dramatically affect the research program that they suggest. The preceding discussion may be simply an in-house quibble among similarly inclined theorists. Nonetheless, it is interesting to speculate about the source of the apparent confusion. The explanation may lie in a continuing tendency to associate the jury control idea with the design of exclusionary rules and to associate the advocate control idea only with the choice between an adversarial and a comparatively inquisitorial system of procedure, as if the advocate control principle drops out of the mix of consideration once the choice is made to use an adversarial system with (limited) rules

---

Emerging Findings, General Issues and Future Directions, 76 Minn. Rev. 65 (1992) (reporting results of mock jury experiments).

[42] See Allen & Leiter, supra note 1, at 1499 (quoting Alvin I. Goldman, Epistemics: The Regulative Theory of Cognition, 75 J. Phil. 509, 510 (1978)).

of pretrial discovery.[43] These unfortunate tendencies persist, even among those who clearly understand the importance of lawyer preparation and presentation of evidence to the development of the common-law exclusionary rules.[44] Perhaps the present discussion will help dissolve these entrenched associations. In an era of continuing assaults on the institution of the jury, however, I think it is unlikely to do so unless we free ourselves from the unhelpful connotations, if not the logical confines, of an emphasis on "epistemic paternalism."

II. EXPECTED UTILITY, RELATIVE PLAUSIBILITY, AND THE BURDEN OF PERSUASION

The setting of the burden of persuasion for a given case allocates the risks of error that remain after all the admitted evidence is considered. Expected utility theory has been applied to explain and to prescribe the criterion of decision if the expected costs associated

---

[43] Such tendencies appear, for example, in the work of philosopher Alvin Goldman, upon whom Allen and Leiter rely. See Alvin I. Goldman, Knowledge in a Social World §§ 9.4–9.7, at 289–304 (1999) (discussing jury control in assessing the veritistic quality of exclusionary rules and discussing lawyer control in assessing the comparative veritism of common-law and civil-law systems of adjudication and the design of discovery rules within the former). To be sure, Goldman correctly recognizes that controlling lawyers' strategic impulses remains an important goal in the adversarial context when formulating rules about expert witnesses. See id. § 9.8, at 304–11. But with regard to the "large class of 'exclusionary' rules" that distinguishes the common-law system, he makes the now rather remarkable assertion that "[t]his peculiarity of the common-law system is worth examining, although it has no obvious relationship to the adversarial character of that system." Id. at 291. While the relationship may not have been obvious at one time, it has been amply demonstrated, see sources cited supra note 23, and is a fact well understood by writers upon whom Goldman himself relies. See Goldman, supra, at 290 (citing Mirjan R. Damaška, Evidence Law Adrift (1997)). Damaška's book begins by noting the competition between jury control and advocate control ideas in regard to the explanation of exclusionary rules. Damaška, supra, at 1–2.

[44] Compare John H. Langbein, The Criminal Trial Before the Lawyers, 45 U. Chi. L. Rev. 263, 300–06 (1978) (observing that evidentiary rules were scarce before lawyers began presenting evidence), and John H. Langbein, Shaping the Eighteenth-Century Criminal Trial: A View from the Ryder Sources, 50 U. Chi. L. Rev. 1, 123–34 (1983) (describing how exclusionary rules developed when lawyers began presenting evidence), with John H. Langbein, The German Advantage in Civil Procedure, 52 U. Chi. L. Rev. 823, 829 (1985) (making the passing comment that the bulk of the exclusionary rules are employed out of a fear of the lay jury's inability to evaluate evidence "purposively").

with these risks of error are to be minimized.[45] "Costs" in this context need not be limited to economic costs in any narrow sense but rather include any negative consequences of erroneous decisions. In cases where a binary choice is presented—guilty or not guilty, liable or not liable—the risks of error are of two kinds. There is the risk of a false positive result, such as convicting the innocent, and the risk of a false negative result, such as acquitting the guilty. If we let $P(G)$ represent the probability that the truth is such that defendant is guilty (or liable), $D(+)$ the disutility of a false positive, and $D(-)$ the disutility of a false negative, then the standard result of expected utility theory is that verdict should be given for the prosecution (or plaintiff) if and only if:

$$P(G) > \frac{1}{1 + D(-)/D(+)}$$ [46]

We should pause to note, in view of the frequent confusion on the matter, that this result does not require the use of Bayes' Theorem, which relates to the question of how to assess $P(G)$.[47] The minimization of expected error costs using such a criterion does not depend on how the decisionmaker arrives at $P(G)$, so long as the assessment is a tolerably accurate expression of the decisionmaker's degree of certainty about the issue after a rational consideration of the admissible evidence, a condition that must be satisfied under any plausible theory of the burden of persuasion.[48]

---

[45] Goals other than cost minimization are also possible within a similarly formalized framework, but error cost minimization is the target of Allen and Leiter's critique. See Allen & Leiter, supra note 1, at 1503–06.

[46] See, e.g., John Kaplan, Decision Theory and the Factfinding Process, 20 Stan. L. Rev. 1065, 1071–74 (1968). For a more general derivation of the decision criterion that takes into account the possibilities of non-zero and unequal utilities for true positives and true negatives, see Dale A. Nance, Evidential Completeness and the Burden of Proof, 49 Hastings L.J. 621, 622–24 (1998).

[47] For a discussion of the role of Bayes' Theorem, see infra Part III.

[48] It is often asserted but never demonstrated that using such a decision criterion *presupposes* perfectly rational decisionmakers properly using Bayes' Theorem to revise a perfectly informed prior probability. See, e.g., Richard S. Bell, Decision Theory and Due Process: A Critique of the Supreme Court's Lawmaking for Burdens of Proof, 78 J. of Crim. L. & Criminology 557, 563–69 (1987). While such conditions would suffice to assure results in accordance with the policy goals reflected in the

Allen and Leiter correctly summarize the implications of this formula:

> In a case involving a binary choice where the disutilities of wrongful verdicts [that is, false positives and false negatives] are equal, decision should be for whomever the probabilities favor. This is the 0.5 rule of civil litigation. . . . If disutilities of wrongful decisions are not equal, as in criminal cases where a wrongful conviction is considerably worse than a wrongful acquittal, the decision rule is adjusted to accommodate the difference.[49]

Allen and Leiter then refer to a number of arguments, ostensibly grounded in veritistic social epistemology, against this way of explaining or prescribing burdens of persuasion.[50] I cannot address all of them here, but I will address enough to demonstrate where I think the alternative approach recommended by Allen and Leiter has value and where I think it falters.[51]

---

decision criterion, there is no reason to assume that they are necessary conditions—in other words, to assume that those goals cannot be achieved to an acceptable degree with more realistic assumptions about how jurors process information to arrive at $P(G)$.

[49] Allen & Leiter, supra note 1, at 1504. Omitted by the ellipsis is this statement: "In cases involving more than two possible explanations, decision should be for the most probable (and here we see the first problem, for this is not the law)." Id. This misstates results in the work of decision theorists, the probable explanation of which is noted below. See infra notes 82, 110, and accompanying text.

[50] Allen & Leiter, supra note 1, at 1504–06. There is a noticeable irony here: Professor Goldman, Allen and Leiter's chief exemplar of modern social epistemology, himself relies on the core insight of expected utility—its emphasis on the comparative utility of false positives and false negatives—in explaining the burden of persuasion. See id. at 1494 & n.9; Goldman, supra note 43, at 284.

[51] In particular, Allen and Leiter mention briefly a set of issues concerning how to account for "base rates," how to conceive of the relationship between policy makers who set the burden of persuasion and judges and juries who apply them, and how to account for problems in the accuracy of probability assessments. See Allen & Leiter, supra note 1, at 1505–06. Space does not permit adequate treatment of these issues here; fortunately, they are carefully analyzed in an article published elsewhere by a defender of expected utility theory. D.H. Kaye, Clarifying the Burden of Persuasion: What Bayesian Decision Rules Do and Do Not Do, 3 Int'l J. of Evidence & Proof 1 (1999).

In place of the expected utility approach, they recommend the "relative plausibility theory",[52] summarized as follows:

> The critical insight of the relative plausibility theory is that legal factfinding involves a determination of the comparative plausibility of the parties' explanations offered at trial rather than a determination of whether discrete elements are found to a specific probability. In civil cases the factfinder is to identify the most plausible account of the relevant events, whereas in criminal cases the prosecution must provide a plausible account of guilt and show that there is no plausible account of innocence.[53]

The relative plausibility idea appears to stand in contrast to the expected utility approach, but the quoted summary reveals something else, a third approach: Determine whether each "element" of the cause of action in question is found to a specific probability and decide for the party with the burden of proof if and only if each of the elements is so found. This approach looks only to the proof of elements of a cause of action or affirmative defense. It is the kind of decision rule Allen and Leiter claim is used by the courts.[54] To sort this out, we must consider the relationship of such elements to the problem of proof.

A. Elements and the Burden of Proof

Every cause of action can be broken down into various components, often called its elements. This breakdown is not unique but serves only to focus attention on particular aspects of the cause of action as a whole. For example, a cause of action for negligence may be broken down into two elements: (A) breach of duty and (B) compensable injury proximately caused by that breach. Alternatively, one can break the same cause of action down into four elements: (W) existence of a duty; (X) breach of that duty; (Y) existence of compensable injury; and (Z) a proximate causal connection between the breach and the injury. The choice of the number of elements does not affect what must be proved to prevail on the cause of

---

[52] This theory was first presented by Allen some fifteen years ago. See Ronald J. Allen, A Reconceptualization of Civil Trials, 66 B.U. L. Rev. 401, 425–31 (1986).

[53] Allen & Leiter, supra note 1, at 1527–28.

[54] Id. at 1504.

action so long as each articulation is true to the substantive law.[55] Assuming away doubts about the facts—that is, assuming one knows the facts with certainty—the plaintiff should prevail if and only if each element is present. Deductively, using a two element breakdown with A and B representing the elements, the conjunctive event $\{A$ and $B\}$ is true if and only if A is true and B is true, which is true if and only if $\{W$ and X and Y and $Z\}$ is true, and so forth.

But when one takes account of the problem that the truth about litigated facts is almost never known with certainty, things get more complicated. In the context of a simple case with no affirmative defenses or alternative claims, the plaintiff should win if and only if all elements of the plaintiff's single cause of action have been adequately proved (without specifying for now what "adequately proved" means). Unfortunately, that cannot be easily translated into a series of statements about the individual elements. One might try to say, using two elements again, that the conjunctive proposition $\{A$ and $B\}$ is proved to sufficient certainty if and only if A is proved to sufficient certainty and B is proved to sufficient certainty. Unless one abandons the traditional axiomatic rules of probability theory, however, the measure of what is "sufficient" in the proof of the conjunctive event $\{A$ and $B\}$ will not be the same as the measure of what is "sufficient" in the proof of A and of B taken serially, as Allen and Leiter observe.[56]

---

[55] To be sure, it might have psychological effects on a jury that are not logically warranted.

[56] See Allen & Leiter, supra note 1, at 1504. This is the supposed *paradox of conjunction*: Just because $P(A) > p$ and $P(B) > p$, one cannot infer that $P(A$ and $B) > p$. For example, if $P(A) = 0.6 > 0.5$, $P(B) = 0.6 > 0.5$, and $P(B|A) = 0.7$, then $P(A$ and $B) = P(A) \times P(B|A) = 0.42 < 0.5$.

To obtain needed precision, let us restate the three theories, as applied in the context of a civil case with a single cause of action, two elements A and B, and no affirmative defenses, using $P(X)$ to mean the probability that X is true and using the conventional "iff" to mean "if and only if." Under the expected utility theory, if the disutility of a false positive is equal to the disutility of a false negative, the decision criterion is:

Criterion 1:   Plaintiff wins iff
(expected utility) $P(A \text{ and } B) > 0.5$

So that this criterion cannot be bested by competitors merely on the ground that it involves explicit quantification, which might be thought antithetical to law or at least to the ability of jurors to comprehend instructions, a mathematically equivalent expression that does not require the computation of a number is:

Criterion 1:   Plaintiff wins iff
(alternative form) $P(A \text{ and } B) > P(\text{not } (A \text{ and } B))$

This criterion prescribes a plaintiff victory if the conjunctive proposition $\{A \text{ and } B\}$ is more probable than not.

Those familiar with probability theory will know that, tautologically, $P(A \text{ and } B) = P(A) \times P(B|A)$. In the event that A and B are stochastically independent, this equation reduces to $P(A \text{ and } B) = P(A) \times P(B)$. It is important to emphasize, however, that nothing about expected utility theory requires the factfinder in a lawsuit to assess the probability of the conjunctive event by using these multiplicative properties after independently assessing $P(A)$ and $P(B)$ (or, in a case of stochastic dependence, $P(B|A)$). Although reference to these properties may be useful in particular trial contexts, the factfinder ordinarily will and presumably should use whatever heuristic devices are suitable or familiar for the type of inferences involved in the trial.[57]

---

[57] Some of the criticisms of expected utility theory that one encounters in the literature fail to acknowledge this point, and a similar failure may motivate Allen and Leiter in their rejection of expected utility theory as among those "algorithmic" approaches that factfinders are practically incapable of using. For a discussion of the use of the probability calculus in thinking about $P(G)$, see infra Part III. Of course,

The relative plausibility theory as advanced by Allen and Leiter is more difficult to state explicitly in comparable terms, owing to ambiguities in the way they have expressed it. One source of ambiguity is the use of the term "plausibility" instead of "probability." This requires one to assess in what respect those terms are different, but Allen and Leiter are silent on this point. From earlier work by Allen, it is clear that plausibility is in part a matter of probability, but the relationship is unclear.[58] In the rest of their paper, Allen and Leiter consistently use comparative probabilities as an index to comparative plausibilities, and so, therefore, will I.[59] With this caveat, the criterion seems to be:

Criterion 2:   Plaintiff wins iff
(relative plausibility) P(plaintiff's story) $>$ P(defendant's story)

Allen and Leiter make much of the "comparative" nature of juridical factfinding and its coherence with the relative plausibility theory, but alternative Criterion 1 is just as "comparative" as the relative plausibility theory. The difference is just that the probabilities being compared are not the same under the two theories. The other principal advantages Allen and Leiter claim for Criterion 2 flow from the fact that the jurors need not attempt to search over all possible stories that would explain the evidence—which may indeed seem like an impossible task—but rather can limit their consideration to the stories (or explanations) advanced by the par-

---

the veritistic tendencies of commonly applied heuristics are much debated. Compare Amos Tversky & Daniel Kahneman, Judgment Under Uncertainty: Heuristics and Biases, *in* Judgment Under Uncertainty: Heuristics and Biases (Daniel Kahneman et al. eds., 1982) (presenting a relatively pessimistic assessment), with Gerd Gigerenzer et al., Simple Heuristics That Make Us Smart (1999) (presenting a more optimistic assessment).

[58] Allen has written that: "What makes stories plausible seems to be determined by such variables as coherence, consistency, completeness, uniqueness, economy, and (yes) probability." Ronald J. Allen, Rationality, Algorithms and Juridical Proof: A Preliminary Inquiry, 1 Int'l J. Evidence & Proof 254, 274 (1997). One would think, for example, that a more complete, coherent, consistent, and economical story is a more probable story, but Allen's explication does not clearly indicate the relationship.

[59] If it is not true that any differences in the plausibility of the contending stories is reflected in their comparative probabilities, then the arguments that follow with regard to the relative plausibility theory are more complicated; I think, however, they remain essentially correct.

ties.[60] The idea, explained more fully in an earlier article by Allen, is that the judge would check the story advanced by each party to assure that, if true, it would require a judgment for the offering party under the substantive law. Only if both parties advance such stories is a trial necessary, with the jury instructed simply to decide in favor of the party whose story is more plausible.[61]

Finally, Allen and Leiter claim that under prevailing legal doctrine, the decision criterion is that the plaintiff should win if and only if each element of the cause of action is proved, seriatim, by a preponderance of the evidence:[62]

Criterion 3:	Plaintiff wins iff
("elements" approach)	$P(A) > 0.5$ and $P(B) > 0.5$

Obviously, this criterion assumes that the civil standard is "more probable than not," either nominally or as an interpretation of the phrase "preponderance of the evidence."[63] Again, to avoid the red herring of explicit quantification, the criterion can be restated in the following non-numerical, comparative form:

Criterion 3:	Plaintiff wins iff
(alternative form)	$P(A) > P(\text{not-}A)$ and $P(B) > P(\text{not-}B)$

With these criteria expressly articulated, one can restate the Allen and Leiter thesis as follows: Both the expected utility theory and the relative plausibility theory diverge from extant doctrine, which is represented by Criterion 3. Extant doctrine is wrongheaded for many reasons, and, again for many reasons, the relative plausibility theory (Criterion 2) is better than the expected utility theory (Criterion 1) as an alternative to extant doctrine.[64] For convenience, it is appropriate to summarize here the response that is detailed in the following sections. First, there is in fact very little

---

[60] Allen & Leiter, supra note 1, at 1527–28.

[61] Ronald J. Allen, The Nature of Juridical Proof, 13 Cardozo L. Rev. 373, 406–13 (1991).

[62] Allen & Leiter, supra note 1, at 1504.

[63] See Allen, supra note 52, at 405.

[64] Allen & Leiter, supra note 1, at 1528 (enumerating the purported advantages of the relative plausibility theory).

difference, if any, between Criterion 1 and prevailing legal doc-trine. Second, Criterion 1 is superior to Criterion 2. Nevertheless, Criterion 2 serves as a useful reference point for the jury in reach-ing a decision under Criterion 1.[65]

B. *Legal Doctrine and The Prescriptions of Expected Utility Theory*

There are certainly good reasons to believe that extant doctrine would be wrongheaded if it were as stated in Criterion 3. In par-ticular, as Allen and Leiter note, such a rule does not minimize expected error costs, and it appears to favor plaintiffs systematically by prescribing plaintiff verdicts in cases where $P(A$ and $B) < 0.5$.[66] In my judgment, however, there is no well-established legal doctrine endorsing Criterion 3 over Criterion 1. The claim that there is such a legal doctrine is most frequently justified—to the extent that it is justified at all—by reference to sample jury instructions.[67] In 1986, I reported my own modest search of authorities and concluded that

---

[65] While Allen and Leiter clearly offer the relative plausibility theory as a *normative* model of decision (one at variance with the prevailing proof rules), they also refer to evidence in the rules and practices of courts as showing that their theory is a more "accurate" description of practice, id. at 1537, which makes it appear that their theory is a descriptive one. Perhaps, therefore, Allen and Leiter are trying to bridge the prescriptive/descriptive gap by offering an "interpretation" of practice with both descriptive and prescriptive components, although they never explicitly say that. If so understood, their claim is simply that, all things considered, Criterion 2 is a better interpretation of prevailing practices than is Criterion 1, proof rules notwithstanding. My reply is that Criterion 1 is a better interpretation of proof rules, even though Criterion 2 may be a useful aid to inferential practices within the framework of such rules.

[66] Allen & Leiter, supra note 1, at 1504. I set aside various attempts to explain why, as a matter of pragmatic accommodation or in recognition of the dynamics of group decisionmaking, using Criterion 3 in practice might better effectuate decisions that actually satisfy Criterion 1 or other considerations of public policy. E.g., Saul Levmore, Conjunction and Aggregation, 99 Mich. L. Rev. 723 (2001); Alex Stein, Of Two Wrongs That Make a Right: Two Paradoxes of the Evidence Law and Their Combined Economic Justification, 79 Tex. L. Rev. 1199 (2001).

[67] See infra note 72. The claim appears to have originated with Jonathan Cohen, a British philosopher, who used it as part of an argument that legal probabilities do not follow the same rules as the probabilities commonly used in science, statistics, decision theory, and so forth. See L. Jonathan Cohen, The Probable and the Provable 58–59 (1977) (stating that "[t]he rule for civil suits requires a plaintiff to prove each element of his case on the balance of probability," and that the conjunction rule, with its multiplicative consequence under mathematical probabilities, "seems to be a rule that is unknown to judges and unrespected by triers of fact"). He provided neither legal authority nor empirical evidence in support of these claims.

standard form jury instructions are usually ambiguous, if not hope-lessly confused, on the point.[68] To my knowledge, no significant legal research since that time has contradicted my conclusion,[69] al-though a number of scholars, including Allen and Leiter, continue to argue from the premise that the law endorses Criterion 3.[70]

In my 1986 article, I also reported a plausible explanation for the confusion that appears in the jury instructions.[71] Suppose that Cri-terion 1 is the test that jurists are clumsily trying to articulate. In that case, it is not surprising that some instructions include state-ments such as, "In order to prevail, the plaintiff must prove every element of his case by a preponderance of the evidence."[72] Even if we construe clauses such as "every element" as denoting a serial consideration of the elements against that standard of proof (which of course is not unavoidable), such instructions are straightfor-wardly consistent with Criterion 1. For example, if $P(A) < 0.5$, then necessarily $P(A$ and $B) < 0.5$, and the plaintiff has failed to prove his case. Consequently, if the jury finds that $P(A) < 0.5$, then it need not go on to consider evidence related to element B. This ef-fects a savings of time and energy for the jury.[73] A serial focus on elements is thus entirely natural and desirable under the expected utility theory. We can incorporate this point directly into the ex-

---

[68] See Dale A. Nance, A Comment on the Supposed Paradoxes of a Mathematical Interpretation of the Logic of Trials, 66 B.U. L. Rev. 947 (1986).

[69] In a recent article, Saul Levmore collects some additional jury instructions and reaches essentially the same conclusion that I did in 1986: that they are varied and often ambiguous in the answer that they give to the question of whether the law employs Criterion 1 or Criterion 3. Levmore, supra note 66, at 724 n.1.

[70] Allen & Leiter, supra note 1, at 1504. For theorists who express strong objection to "rootless theorizing" based on inaccurate doctrinal assumptions, this is no small difficulty. Cf. id. at 1521–27 (criticizing Posner's analyses as "rootless" in that, inter alia, they proceed from false assumptions about the state of evidence doctrine).

[71] Nance, supra note 68, at 949–52.

[72] Typical is the jury instruction relied upon by Allen, supra note 52, at 405 n.19:
    The burden is on the plaintiff in a civil action, such as this, to prove every essential element of his claim by a preponderance of the evidence. If the proof should fail to establish any essential element of plaintiff's claim by a preponderance of the evidence in the case, the jury should find for the defendant.
E. Devitt & C. Blackmar, Federal Jury Practice and Instructions § 71.14 (3d ed. 1977).

[73] For this purpose it is not necessary that jurors *understand* that if $P(A) < 0.5$, then necessarily $P(A$ and $B) < 0.5$.

plicit decision criterion by restating Criterion 1 in the following
logically equivalent form:

| Criterion 1:
(augmented) | Plaintiff wins iff
$P(A \text{ and } B) > P(\text{not } (A \text{ and } B))$, and
$P(A) > P(\text{not-}A)$, and $P(B) > P(\text{not-}B)$ |

This, it seems to me, is a formal expression of what courts are often
trying to articulate, sometimes quite successfully.[74]
    To be sure, a real potential for conflict with Criterion 1 can arise
if the jury is instructed, in effect, with the *converse* of the above in-
struction: "Plaintiff prevails if the plaintiff proves each element of
his case by a preponderance of the evidence." It is easy to see how
lawyers and judges, betrayed by their lack of sophistication in how
probabilities work, might slide into the latter element-by-element
instruction from the former, even though the two formulations are
not logically equivalent. *If* the jury understands the latter instruc-
tion to call *only* for a serial testing of elements by the more likely
than not standard, and *if* the jury follows this instruction, then they
might give verdict for the plaintiff in cases for which, under Crite-
rion 1, they should not.[75] It is difficult, however, to find jury
instructions that are clearly of this type.[76] Moreover, even if juries
were given such an instruction, it is entirely possible that they

---

[74] Consider, for example, the following instruction:
      For the plaintiff to recover from the defendant on his claim of negligence, you
      must find that all of the following have been proved by a preponderance of the
      evidence: 1. The plaintiff had injuries; 2. The defendant was negligent; and 3.
      The defendant's negligence was a cause of the plaintiff's injuries. If you find
      that any one or more of these statements has not been proved, then your
      verdict must be for the defendant. On the other hand, if you find that all of
      these three statements have been proved, then your verdict must be for the
      plaintiff.
Colo. Supreme Ct. Comm'n on Civil Jury Instructions, Colorado Jury Instructions
§ 9.1 (4th ed. 1999).
[75] For an illustration, see supra note 56.
[76] Judgments entered pursuant to special verdict forms that require the jury to
answer only questions about the proof of individual elements may produce such
results. To be sure, that would not necessarily mean that the difference between
Criterion 1 and Criterion 3 is appreciated when such judgments are entered. See
David A. Lombardero, Do Special Verdicts Improve the Structure of Jury Decision-
Making?, 36 Jurimetrics J. 275 (1996) (examining the conjunction problems created
by special verdicts).

would understand it as consistent with Criterion 1 in any case where the difference between Criterion 3 and Criterion 1 would in fact matter. For example, if "preponderance of the evidence" is not construed as measured by an invariant specific probability, the jury might implicitly construe that standard as applied to a specific element in such a way as to assure that the verdict will be for plaintiff only if $P(A$ and $B) > 0.5$. In other words, "preponderance of the evidence" might be taken to mean one thing as applied to a single element and another as applied to the conjunction of all elements.[77]

In the final analysis, what little difference there may be between these prescriptions of expected utility theory and extant judicial utterances about the burden of persuasion may be largely or wholly unintended, as well as ineffectual, in terms of regulating the conduct of jurors. Beyond that, one need not rest entirely on the examination of jury instructions and other products of judicial reasoning for the conclusion that expected utility theory is closely related to doctrine and practice. Powerful evidence exists in constitutional due process jurisprudence. Numerous decisions by the United States Supreme Court have employed expected utility theory, more or less explicitly, to analyze arguments that a given burden of persuasion, specified by statute or common law, violates the constitutional guaranty. Conspicuous in the opinions in such cases, whether majorities or dissents, is the common-sense assessment of the relative costs of false positives and false negatives, just as expected utility theory suggests.[78] Of course, this jurisprudence can be criticized,[79] but that does not change the fact that the courts

---

[77] Following a suggestion by Professor John Kaplan, I have previously argued that the very imprecision of the verbal formulae in which burdens of persuasion are stated have the virtue of allowing the jury to participate at the margins in setting the level of the burden by their interpretation of the standards. See Nance, supra note 46, at 624. Here, it is suggested that the jury might use this flexibility to avoid probabilistically incoherent results.

[78] See, e.g., Rivera v. Minnich, 483 U.S. 574, 579–81 (1987); id. at 583–85 (Brennan, J., dissenting); Santosky v. Kramer, 455 U.S. 745, 758 (1982); id. at 186–89 (Rehnquist, J., dissenting); Addington v. Texas, 441 U.S. 418, 425–27 (1979).

[79] See, e.g., Michael L. DeKay, The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof, 21 Law & Soc. Inquiry 95 (1996) (arguing that minimization of *expected* error costs using an appropriate relative weighting of errors, though the better policy, should not be confused, as it sometimes is, with minimization of *actual* error costs using the same weighting).

have instantiated expected utility theory in reasoning about the burden of persuasion.[80]

It is therefore quite astonishing that Allen and Leiter assert that there is "virtually no . . . evidence" for the expected utility theory in the case law.[81] This is not to say that extant doctrine never diverges from the prescriptions that expected utility theorists have generated. In particular, the law relating to cases with multiple defendants or multiple plaintiffs is complex and may not correspond in all contexts with such theoretical prescriptions.[82] Much of this theory and law is also relatively new and evolving, however, and it is hard to know how much they will diverge from one another once they mature and settle into some quasi-stable equilibrium. In any event, it is very difficult to deny the usefulness of expected utility theory in thinking about proof burdens.

C. A Close Look at the Relative Plausibility Theory

There is no plausible interpretation or reconstruction of the usual jury instructions regarding burdens of proof that can cause a convergence of doctrine with the relative plausibility theory,[83] and this remains true if that doctrine is closer to Criterion 1 than to Criterion 3. The main reason is that the relative plausibility theory is stated in terms of *stories advanced by the parties*; indeed, this is the source of its claimed advantages. Yet it also entails the possibility of a story that would explain the evidence presented but that is not offered by either party, which in turn presents the potential for serious divergence between Criterion 2 and each of the other criteria. In my view, this "third story" possibility is the most serious problem for the relative plausibility theory, one that needs to be

---

[80] The seminal opinion is that of Justice John Marshall Harlan, who explicitly cites an important early article using expected utility theory. In Re Winship, 397 U.S. 358, 370 n.2 (1969) (Harlan, J., concurring) (citing Kaplan, supra note 46).

[81] Allen & Leiter, supra note 1, at 1537.

[82] See, e.g., Daniel A. Farber, Toxic Causation, 71 Minn. L. Rev. 1219, 1238–57 (1987) (analyzing the multiple plaintiff problem); David Kaye, The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation, 1982 Am. B. Found. Res. J. 487, 503–08 (analyzing the multiple defendant problem).

[83] See Allen & Leiter, supra note 1, at 1537.

addressed adequately before it is embraced.[84] The crucial question is this: Have Allen and Leiter advanced the discussion of this problem in their present paper?

In an interesting and revealing passage, they reply as follows to an analysis by Posner concerning this issue:

[Posner] asserts that if

> the plaintiff's story had a probability of .42 of being true, the defendant's story a probability of .30 of being true, and the probability that another story or stories is true was .28, then the plaintiff should lose because he has failed to prove that his story is more likely than not true.

. . .

[Allen and Leiter reply:] One either knows or does not know the implications of the story or set of stories comprising the missing 0.28 probability. If these implications are known, each party should get the benefit of the probability associated with the story or stories that favor them. If the implications are not

---

[84] The problem was identified in commentaries on Allen's earlier articles. See Richard Lempert, The New Evidence Scholarship: Analyzing the Process of Proof, 66 B.U. L. Rev. 439, 471–77 (1986); John Leubsdorf, Stories and Numbers, 13 Cardozo L. Rev. 455, 458–59 (1991). To be sure, there are other problems with the relative plausibility theory, some of which are addressed in the above-cited articles. One difficulty that these articles do not address is the fact that the relative plausibility theory, at least as so far presented, aggravates a problem that already afflicts the usual account of the conventional proof rules, whether those rules are understood (in the civil context) as Criterion 1 or as Criterion 3. That is the problem of missing evidence, or as I have called it, "evidential incompleteness." See Nance, supra note 46, at 626–32. Without further constraints, either Criterion 1 or Criterion 2 might be satisfied in cases where reasonably available evidence is not presented in court. This possibility becomes more pressing when verdicts for one side are allowed even though the probability of that party's story being true is less than 0.5, because one response of jurors to the absence of such evidence might well be to discount the probability of either or both of the parties stories. For example, faced with very little of the available evidence relevant to a case, the jury might assess P(plaintiff's story) = 0.2 and P(defendant's story) = 0.1, and under Criterion 2, plaintiff should prevail. This is problematic because the law's coercive mechanisms should not be made available to private parties under such a paucity of evidence. The solution, under either theory, is to incorporate a further constraint under the rubric of the burden of production. See id. at 625–26.

known, there is no good reason to systematically disfavor plaintiffs by attributing all the ambiguity to them.[85]

Posner is wrong, to be sure, and for essentially the reasons stated by Allen and Leiter. But the relative plausibility theory does *not* prescribe the result that Allen and Leiter endorse for this situation. Criterion 2 clearly provides that plaintiff wins even if the third story, garnering the 0.28 probability, favors defendant, because *by hypothesis* that story is not the story that defendant presents.

Allen and Leiter have an answer of sorts to this problem. They continue their response to Posner by arguing, "[i]n civil cases, given mutual discovery, the parties can be expected to search for and produce evidence of whatever stories they think can plausibly support their legal claims."[86] This, of course, changes the hypothetical by assuming that if the third story favors the defendant, the defendant will have advanced that story as an alternative theory of the case.[87] Allen and Leiter's response must be construed as an assertion that hypotheticals like the one posited by Posner will not occur so long as at least one side of the dispute can discern that the third story would be favorable to that side under the substantive law.[88] But is this true?

---

[85] Allen & Leiter, supra note 1, at 1530–31 (quoting Posner, supra note 12, at 1513) (footnote omitted).

[86] Allen & Leiter, supra note 1, at 1531.

[87] One might object that the whole point of the relative plausibility theory is undermined by allowing alternative stories, a step that begins to move the whole scheme in the direction of its rival. If the primary virtue of the theory, however, is sufficient simplicity to permit juries to adhere to its prescriptions, then it is not fatally flawed by permitting a *limited* number of alternative stories to be advanced by a party.

[88] Ironically, a similar proposition was advanced by Professor Richard Lempert in *criticizing* the relative plausibility theory. His argument was that there will be few situations in which the relative plausibility criterion of decision will in fact diverge from the "more likely than not" criterion. See Lempert, supra note 84, at 473–74 (arguing that in nearly all cases a "spoliation inference" will cause the jury to reevaluate the probabilities so as to rule out the third story, reasoning that if there were evidence supporting the third story, the party favored by it would have offered that evidence). For reasons explained below, I think Lempert is wrong on this point; there are often cases in which no such spoliation inference should or will occur. In any event, Lempert does acknowledge at least the possibility that cases will arise with significant probabilities assigned by the jury to stories not advanced by either party, and in that eventuality Lempert favors the solution prescribed by Criterion 1. Id. at 474–75.

Consider what I will call "the traffic signal problem." Plaintiff alleges that defendant ran a red light, causing an accident. Defendant alleges that his light was green. At trial, the parties tell their stories as alleged. There is only one element, negligence that the parties dispute, and the color of the light is the only material fact determining negligence. The jury assesses the probabilities as follows:

P(defendant entered on *red* light) = 0.42
P(defendant entered on *green* light) = 0.30
P(defendant entered on *yellow* light) = 0.28

Notice that, regardless of which party the law favors if the light was yellow, each litigant has successfully settled on the theory of the case that the jury finds most plausible among those that favor that litigant. Criterion 2 prescribes that plaintiff should win because 0.42 > 0.30, and this is so even if the rule of law is that entering on a yellow light is not negligent.[89] According to Allen and Leiter, however, in that event the 0.28 probability should inure to the benefit of defendant, and plaintiff should lose because 0.42 < 0.30 + 0.28. Note that this is the result under Criterion 1, the rule prescribed by the expected utility theory. Presumably, Allen and Leiter would try to escape this dilemma by arguing, as they do in response to Posner, that the hypothetical is wrong and that in reality such a defendant would plead and testify in the alternative, telling the story that the light was either green or yellow. This would generate the same defense verdict, but without having to rely on Criterion 1 since Criterion 2 gives the same result.

Does this strategy work? For theorists concerned with realistic behavioral assessments, it is striking that Allen and Leiter would reject out of hand the possibility that the defendant, wanting to tell a single coherent story, will tell only the story that his light was green. For example, he may believe that by telling a story in the alternative he will present himself to the jury as someone who is unsure of his recollections, thus quite possibly reducing the jury's

---

[89] I am assuming—at this point—that the color of the light *unequivocally* determines whether or not defendant was negligent. Reality is more complicated than that, especially in regard to a yellow light, but the assumption makes it easier to illustrate the point of the hypothetical. The assumption is relaxed in the discussion infra, at notes 95–105 and accompanying text.

assessment of the probabilities of *both* his stories and raising the jury's assessment of the plaintiff's story, which of course may be told with absolute confidence.[90] This behavioral assumption strikes me as reflecting exactly the kind of strategy that clients will be inclined to follow and that advocates will be inclined to encourage in such a case.[91] Beyond that, procedural rules might actually *preclude* the use of alternative stories in particular cases.[92]

The disincentive to tell the yellow light story can be even more pronounced if we change the hypothetical by assuming, more realistically, that the rule of law is that entry on a yellow light may or may not be negligent depending on how the circumstances are evaluated under a general "reasonableness" standard. In such a case, it is quite possible that neither party will want to tell the story that the light was yellow, each one fearing that such a claim would be taken as a concession of doubt and also worrying that the jury might resolve the judgmental issue for a yellow light in favor of the opposing party. For risk-averse litigants, and those who just miscalculate, the polarizing tendency in such a context can be powerful indeed.[93]

---

[90] Cf. Leubsdorf, supra note 84, at 459 ("Strategic reasoning might . . . lead a party to avoid a more plausible but more moderate story for fear that the jury will interpret any concession as evidence that there must be a lot more to concede.").

[91] After discovery is completed, clients are often advised that they must eliminate alternative inconsistent pleadings by the time they get to trial so that they can go before the jury with a single unambiguous story. See, e.g., Robert H. Klonoff & Paul L. Colby, Sponsorship Strategy: Evidentiary Tactics for Winning Jury Trials 50 (1990) (advising against going to trial with multiple inconsistent theories of a case).

[92] Allen and Leiter themselves rely on the case of *McCormick v. Kopmann*, 161 N.E.2d 720 (Ill. App. 1959), which concerns the use of alternative pleading and stories at trial. Allen & Leiter, supra note 1, at 1530. The opinion in that case warns that when the pleading party has personal knowledge of which of two incompatible stories is correct, telling both in the alternative is not permitted. *McCormick*, 161 N.E.2d at 727–28. Of course, Allen and Leiter might want to change this procedural rule, but that would still leave the other problems discussed in the text.

[93] In the quoted reply to Posner, Allen and Leiter suggest that cases might arise in which "[o]ne . . . does not know the implications of the story or set of stories comprising the missing 0.28 probability," and argue that in such a case "there is no good reason to systematically disfavor plaintiffs by attributing all the ambiguity to them," as is required by Criterion 1. Allen & Leiter, supra note 1, at 1530–31. Presumably, this refers not to a case like that presented in the text, in which the jurors are able to determine the legal implication of the third story even though the parties cannot predict what the jury will do, but rather to a case in which the jury cannot make the legal determination for such a story. In other words, the jury in such a case

This fact is not lost on juries. Juries, I suspect, commonly believe that neither story told by the parties is very likely to be completely true, that more likely there is some third story that neither party tells exactly—often a compromise taken in part from the plaintiff's story and in part from the defendant's.[94] Indeed, the social science upon which Allen and Leiter rely describes jurors as constructing a story from the evidence presented, not passively accepting or rejecting the stories told by the parties,[95] and there is no reason at all to believe that such a constructed story will necessarily or even likely match one of the stories told by the parties, certainly not in every particular. Under the expected utility theory and the interpretation of prevailing doctrine that is compatible with it, the jury must consider which side is favored legally by that compromise story and attribute its assessed probability accordingly.[96] This will not happen under the relative plausibility model, though, unless ju-

---

cannot say whether the story to which they attribute a 0.28 probability is one that, if true, favors the plaintiff. This appears to be simply a defect in the instructions given on the substantive law and would call for clarification. Absent such clarification, in a case for which the missing 0.28 probability would make the difference between a verdict for the plaintiff and a verdict for the defense, I can see no reason not to treat the case the same as one would treat those unusual cases in which, on purely factual grounds, the jury has no confidence that the decision criterion has been met—the "equipoise" cases for which the law assigns the verdict to the defendant by default. The "good reasons" to do so are the same: institutional inertia to avoid enforcement costs and (when applicable) to avoid the stigma of fault attributions about which the system does not have confidence.

[94] Cf. Klonoff & Colby, supra note 91 (developing at book length the implications of the proposition that juries assume the truth can be *no more* favorable to a party than the party's assertions).

[95] See Nancy Pennington & Reid Hastie, A Cognitive Theory of Juror Decision Making: The Story Model, 13 Cardozo L. Rev. 519 (1991), cited in Allen & Leiter, supra note 1, at 1528 n.110.

[96] Professor David Kaye illustrates this in his effort to formalize the use of "stories" in expected utility theory terms. He expresses the probability of interest as "$\Pr(S^p|E)$," meaning the probability that plaintiff's story is true given the evidence in the case, David H. Kaye, Comment, Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?, 66 B.U. L. Rev. 657, 661 (1986), and argues, in the context of criminal cases:

> Some factual contentions in S^p may not be essential to satisfying the elements of the offense. For instance, the prosecution may argue that the defendant acted with a particular motive, but the jury may convict even though it concludes that the defendant acted for a different reason. Perhaps S^p should be thought of as the minimal body of contentions along the lines suggested by the prosecution that, if believed, would warrant a verdict of guilty.

Id. at 661 n.10.

rors ignore the instructions that would have to be given pursuant to it.

For example, as a modification of the traffic signal problem, suppose the rule is that entry on a yellow light must be evaluated under a general reasonableness standard, while entry on red or green is still governed by a *per se* rule. Once again, the plaintiff testifies that defendant's light was red, but now the defendant testifies that his light was either green or yellow and that—in order to address the reasonableness standard—defendant could not see the plaintiff's car enter the intersection to make a turn after a stop on a red light. Suppose further that the jury considers the "I couldn't see him" claim as likely to be true as its negation, resulting in the following attributions of probability:[97]

Color of defendant's light	Defendant could NOT see plaintiff	Defendant could see plaintiff	Total probability
Red	0.21	0.21	0.42
Green	0.15	0.15	0.30
Yellow	0.14	0.14	0.28

Shading indicates those three of the six basic stories that, if known to be true, would require a verdict for the plaintiff under the assumed substantive law; the stories for the three unshaded cells favor the defendant. In such a case, should the jury give defendant the benefit of the $0.30 + 0.14 = 0.44$ probability that the light was either green (no negligence) or yellow-without-seeing-plaintiff (no negligence in this context), or should it only credit defendant with the $0.15 + 0.14 = 0.29$ probability that the light was either green-without-seeing-plaintiff or yellow-without-seeing-plaintiff? If the former, defendant wins, because 0.42 is less than 0.44. If the latter, plaintiff wins, because 0.42 is greater than 0.29. To be faithful to the premises of the relative plausibility theory, the answer would have to be the latter, even though the "I couldn't see him" claim is

---

[97] The table assumes for simplicity that the probability that the defendant could see the plaintiff is independent of the light color and that the choice of story that defendant tells has not affected the total probabilities associated with each light color when compared to the problem as originally presented.

immaterial under the hypothesis of a green light. The jury has no flexibility to attribute to the defendant the "compromise" story that the light was green but the defendant could see the plaintiff's car because that is not a story the defendant presented.[98]

Once again, parties might try to avoid this kind of problem by telling alternative stories. For example, the defendant in this example might advance the following claims: (a) his light was green and he could not see the plaintiff's car; (b) his light was green even though he could see the plaintiff's car; or (c) his light was yellow and he could not see the plaintiff's car. Just to articulate this possibility reaffirms the point made above, that parties may be unwilling (or not permitted) to make such allegations expressly or to advance such explanations in the evidence presented. The problem only becomes worse as additional pieces of the parties' stories are rejected by the jury or even just considered less than certainly true.[99]

In the end, the relative plausibility theory will acceptably handle the problem of the third story only under the remarkably fortuitous condition that the attribution to the parties of probabilities associated with third stories would not affect, in any significant number of cases, the results of the probabilistic comparison of the stories that *are* advanced by the parties.[100] Maybe this condition holds, but there is no empirical evidence that it does. A priori considerations do not lead to any obvious conclusion on the matter except that it is very problematic. Most importantly, the polarizing effect described above may well create situations in which the most plausible story is not told by either party, and the effect of attributing the most plausible story could well swamp the direct plausibility comparison of the stories that *are* told.[101] There are also more subtle

---

[98] The particular example presented takes advantage of the applicability of both a per se negligence rule and a general reasonableness standard in the context of the same case, but other examples can easily be constructed that do not have this feature.

[99] Obviously, one might stipulate that the jury is free to "modify" the parties' stories as it believes the evidence warrants and then to compare the stories as modified. But this is simply to begin the process of moving from Criterion 2 to Criterion 1, and it is very hard to see any coherent stopping point on that progression.

[100] At one point, Allen appeared to believe that this would be the case, though he gave no reason to support that belief. See Allen, supra note 61, at 410 n.118.

[101] It is reported that the physicist E.T. Jaynes has shown that "a good approximation of the probability of an hypothesis can usually be attained by comparing it with the next most

effects that make satisfaction of the indicated condition problematic.

For example, in the original traffic signal problem, for which a yellow light unequivocally favors defendant but neither party tells the story that it was yellow, attribution of the probability assigned to the yellow light story—though not the most plausible of the three—still changes the result that the relative plausibility theory produces. Indeed, that was the whole point of the example. In the perhaps more realistic modification of that hypothetical, where a yellow light throws things into a judgmental "reasonableness" standard and the issue arises whether defendant could see the plaintiff's car, the question whether such attribution affects the result under the relative plausibility theory depends fortuitously on exactly which stories are advanced by the parties. If, on the one hand, the plaintiff asserts that the defendant's light was red and the defendant could see his car (probability of 0.21), and the defendant asserts that the light was green and he could not see the plaintiff's car (probability of 0.15), then attribution of the probabilities associated with other possibilities does not affect the result. On the other hand, if the plaintiff tells the story that the light was red and the defendant could see him (probability of 0.21), while the defendant asserts that the light was either green or yellow and he could not see the plaintiff (probability of 0.29), then attribution of the other probabilities does affect the result. A number of other combinations can be imagined. In other words, a great deal—too much—turns on the strategic decisions of the parties about which stories to endorse in their pleadings and testimony.

One further, but related, set of problems should be noted. In previous papers, Allen has argued that the jury need not be instructed on the elements of the substantive law at all since the judge will have performed the check to see if the stories advanced satisfy the substantive law requirements applicable to each side.[102]

---

likely hypothesis." Bernard Robertson & G.A. Vignaux, Probability—The Logic of the Law, 13 Oxford J. Legal Stud. 457, 471–72 (1993). Putting aside the question of how good of an approximation is good enough, such a result obviously will not help the relative plausibility theory if the parties do not tell the two most likely stories. Id. at 472.

[102] See Allen, supra note 61, at 410 n.118. In the same passage, Allen allowed for the possibility that jurors might be instructed in the elements of a cause of action and then

This would appear to be a significant advantage, eliminating what many consider to be cumbersome and poorly understood jury instructions on the substantive law. There is a significant price to be paid for this advantage, however. First, putting aside problems of judgmental standards that the jury is supposed to apply but about which they will be unaware absent instruction, and assuming once again that there is a determinate legal rule for entry on a yellow light, a lack of substantive instruction would complicate the jury's task considerably. For example, the jury would not have the benefit of an instruction on the significance of a yellow light under the substantive law, and so could not attribute the probability associated with the yellow light story to either party except by legislating for such cases. It may not be bad, all things considered, for juries to make their own law in some cases, and the use of judgmental standards like "reasonableness" often has this effect, but why should the jury's authority to do so depend on the fortuity of whether or not the parties have advanced the story to which some law must be applied?[103]

Second, without instructions on the substantive law, the jury cannot know which facts the law regards as material. Consequently, it cannot distinguish those parts of a party's story that can be disregarded in assessing relative plausibility. For example, if plaintiff's story in the traffic signal problem included a reference to some immaterial fact, such as the color of the hat he was wearing that day, that the jury comes to believe is completely implausible, may it disregard that factual assertion in assessing the plausibility of the plaintiff's story as a whole?[104] Again, the jury could be left to its own devices in deciding whether to disregard that matter, and the jury's common-sense judgment would undoubtedly be consistent with the law's judgment in some cases but not in others. Even in those cases in which it is consistent, that degree of happy fortuity

---

permitted to apply them to third stories. Id. In that case, however, his relative plausibility theory unravels *as a decision criterion.* See infra Section II.D.

[103] With regard to the situation in which the jury would believe, rightly or wrongly, that they have no authority to make up the law when none is given to them, see supra note 93.

[104] Of course, the inaccuracy of the plaintiff's story in regard to the color of the hat might have indirect effects on the material facts by influencing plaintiff's credibility. But even if the plaintiff loses credibility, the probability of the material portions of the plaintiff's story being true might still be much higher than the probability that plaintiff's entire story is true.

depends on the jury's being aware that it has the authority to make such judgments at all; jurors might well believe that they must accept either the plaintiff's story *as is* or the defendant's story *as is*. Indeed, this might be true even if the jury *is* instructed on the substantive law if it is also instructed in accord with the relative plausibility theory. In order to avoid such a result, any instruction based on Criterion 2 would need to be combined with instructions on the substantive law *and* an instruction that the jury may disregard immaterial portions of the stories told, except insofar as they relate to credibility. Like other changes that would be needed to make the relative plausibility theory work acceptably, this would cause that theory to converge toward the theory it was designed to displace.

D. The Most Plausible Story Theory

As already observed, the empirical literature has emphasized the jury's role in reconstructing what happened out of the evidence that the parties present. This suggests a rather different theory that sometimes appears in the arguments. Recall Allen and Leiter's summary regarding civil cases:

> The critical insight of the relative plausibility theory is that legal factfinding involves a determination of the comparative plausibility of the parties' explanations offered at trial rather than a determination of whether discrete elements are found to a specific probability. In civil cases the factfinder is to identify the *most plausible* account of the relevant events . . . .[105]

If one reads the second sentence as unmodified by the first, then the use of the superlative term "most" in the second sentence, rather than the comparative "more," suggests that the jury's task would be to identify the single most plausible story, whether or not advanced by a party, presumably then applying the substantive law—about which it would have to be instructed—to that most plausible (probable) story by determining whether or not it satisfies all the elements of the cause of action.[106] This theory invites us

---

[105] Allen & Leiter, supra note 1, at 1527–28 (emphasis added).

[106] The alternative to instructing the jury on the elements of the substantive law would be to require the jury to report the most plausible story in detail to the judge, who would then apply the substantive law. This would be a form of a special verdict.

to conceptualize the reconstruction in terms of the jury generating a set of mutually incompatible, plausible accounts that are, or at least could be, rank ordered from most plausible to least plausible. The jury then selects the one at the top of the list: the most plausible account.[107]

This "pick the most plausible story (whether advanced by a party or not)" theory is an interesting one in itself. In some ways, it seems analogous to the "pick the best heuristic" approach to decisionmaking, which in various contexts has been shown to produce surprisingly accurate decisions.[108] According to the latter theory, a decisionmaker focuses on just one important factor in making a decision, ignoring all others. That theory, however, is considerably different from the most plausible story theory, which does not isolate a specific evidential factor in a case (such as the credibility of the sole witness to an event) but rather isolates the single most plausible story that would explain the evidence. Whether that approach, applied across the long run of trials, would enjoy accuracy and economy of cognitive resources at levels that would be acceptable in the legal context is, once again, an open and largely empirical question. There is reason, however, to be skeptical. In particular, the most plausible story approach suffers from problems similar to those besetting the relative plausibility theory.

Applied, for example, to the traffic signal problem, the "most plausible story" approach would require a verdict for the plaintiff even if the jury believes that it is more likely than not that the light was either yellow or green and thus that defendant was not negligent.[109] By definition, under this theory, there could be no aggregation of probabilities associated with distinct and incompatible stories.

Such an awkward procedure would strongly encourage challenges to the jury's verdict based on insignificant details or infelicities of expression in the jury's report.

[107] Those accounts not at the top of the list might not need to be fully worked out so long as the jury has confidence that none can be worked out in such a fashion as to move into first place. The possibility that no single story is most plausible could be handled easily enough, at least in the most probable context of such an unlikely event, a two-way tie for the most plausible. Either both stories favor the same party and the decision is easy or they split between favoring the plaintiff and favoring the defendant, in which case the familiar default rule in favor of the defendant would presumably apply.

[108] See Gigerenzer et al., supra note 57, at 73–188.

[109] In terms of the previous discussion, this assumes either that the law attributes non-negligence to an entry on yellow or that the jury determines such an entry to be non-negligent under the particular circumstances.

Even if the defendant presented the case in the alternative, arguing that the light was either green or yellow, the jury would be required to accept the single most plausible story—that the light was red. If disjunctive stories could be considered as a single story, the whole idea collapses. Specifically, if *any* mutually incompatible stories could be combined to form a "single" disjunctive or alternative story, then the single most plausible story would always be the disjunction of all possible stories, with a probability of 1. If only those mutually incompatible stories favoring the same party could be combined into a disjunctive story, then the theory collapses into Criterion 1. I suspect that the latter is what Allen and Leiter had in mind, and their reply to Posner's three story hypothetical tends to confirm this, for they do not argue that the plaintiff in Posner's hypothetical should win just because his explanation is the most plausible of the three.[110]

The most plausible story approach can safely ignore aggregation problems only on the condition that the attribution of probabilities associated with all stories that are *not* the most plausible would not affect the verdict dictated by the most plausible story in any significant number of cases. That means downplaying the traffic signal problem (in various permutations) as atypical, not on the ground that one of the parties will always tell the yellow light story as an alternative, but on the distinct ground that only in a small percentage of cases, small enough to ignore for practical purposes, will the

---

[110] See supra text accompanying note 85. Allen and Leiter's occasional use of language suggesting a most plausible story theory may be motivated by a desire to account for the rules that arguably should apply in some, but not all, cases with more than two parties, rather than two party cases with more than two stories. See supra note 82 and accompanying text. At one point, Allen and Leiter discuss the three party case of *McCormick v. Kopmann*, 161 N.E.2d 720 (Ill. App. 1959), and make the following assertion: "The jury was essentially instructed to return a verdict against the party—plaintiff or either defendant—most likely liable for the event, just as the relative plausibility theory would predict." Allen & Leiter, supra note 1, at 1530 (footnote omitted). I have read the opinion carefully and can find nothing supporting such a proposition; the only quoted jury instruction concerning the issue of contributory negligence as a defense to the plaintiff's claim against one of the defendants is not written in such comparative terms at all:

> [I]f you find from all of the evidence in the case that [plaintiff] (McCormick) was operating his automobile while intoxicated and that such intoxication, if any, contributed proximately to cause the collision in question, then in that case * * * you should find the defendant, Lorence Kopmann, not guilty.

McCormick, 161 N.E.2d at 725. That reads pretty squarely as Criterion 1.

most probable story be outweighed by less probable stories that cumulate on balance to favor the other side. That condition is perhaps intuitively more plausible than the analogous condition that must be satisfied by the relative plausibility theory, if only because in many cases the most likely story may not be the one told by either party.[111] Nevertheless, the satisfaction of the indicated condition for the most plausible story heuristic also remains both undemonstrated and highly conjectural. As with the relative plausibility theory, it would be unwise to adopt the most plausible story theory as a decision criterion until these issues are adequately addressed.

E. A Synthesis: Distinguishing Decision Rules from Inferential Methods

Despite what I have argued, I think there is considerable merit in the relative plausibility and most plausible story theories, and I want to state clearly in what respect that is so. To see their real value, one must distinguish between *decision rules* and *inferential methods*. In terms of decision rules, the problem of third stories convinces me, for the time being, that Criterion 1 is superior to both Criterion 2 and its "most likely story" variation, at least for the two party, binary choice case. Nevertheless, most theorists now understand (if ever they did not) that litigants try to tell a coherent story, and—as an inferential rule of thumb—the jury ordinarily should and probably does start by comparing the relative plausibilities of the stories told by the parties.[112] That, it seems to me, is the core of good sense in the relative plausibility idea as well as the import of the various anecdotal statements in case law that Allen and Leiter amass in support of it.[113] If no combination of third stories, whether wholly distinct from those of the parties or in compromise of them, is plausible enough to make a difference in

---

[111] See supra note 100 and accompanying text. As we have seen, the relative plausibility theory severely constrains the jury's freedom to use its epistemic resources in what the jury considers the most effective way by ruling out of bounds all third stories that the parties have not advanced and, implicitly, the heuristic strategies or other considerations that might lead the jury to such stories. The most plausible story theory at least preserves the jury's epistemic freedom to construct a story not told by the parties and so is much less affected by the strategic choices of the parties.

[112] Others have made similar suggestions. E.g., Lempert, supra note 84, at 473; Posner, supra note 12, at 1513; Robertson & Vignaux, supra note 101, at 470–73.

[113] See Allen & Leiter, supra note 1, at 1528–34.

the case, then the jury is done; it need only check the one party's more plausible story against the elements of the substantive law. But if such third stories—including any most plausible story that the jury can construct—make a difference in the result, then the jury must take them into account as well. Thus, aspects of the relative plausibility and most plausible story theories can complement rather than contradict expected utility theory, giving coherence to the former and descriptive depth to the latter.[114]

An obvious advantage of such an accommodation is that it would help to defuse certain puzzling features of the focus on elements of a cause of action. Elsewhere, for example, Allen has argued that such a focus "conflates formal elements and facts":

> Having found the facts, the law is applied deductively for the most part. Thus, the question is not: 'Has each of the elements been proven by a preponderance of the evidence?' The question is instead: 'Given the facts as we have found them, do they entail each of the elements?'[115]

Expected utility theorists understand well that the expression "the probability of negligence" is a shorthand way of saying, "the probability that the events that occurred instantiate the elements of negligence." And judges and jurors surely would see no difference between an instruction that said, "Plaintiff must prove negligence by a preponderance of the evidence," and one that said, "Plaintiff must prove by a preponderance of the evidence events that constitute negligence." Nonetheless, explicit attention to the underlying stories or explanations makes clear that we are thinking about events rather than elements as such, even though our assessment of those events should not be artificially constrained to the accounts advanced by the parties.

Understanding that point, at least implicitly, expected utility theorists tend not to take seriously supposed claims about paradox that are thought to attend Criterion 1 when coupled with the multiplicative property that might be used (but certainly need not

---

[114] It should go without saying that in a particular case the jury might be right not to consider this inferential method to be the one best suited to the task, and I can see no reason to require them to use it. Any attempt to formalize the relative plausibility theory, for example by way of jury instructions, would have to be quite clear that the comparison of plausibilities is merely a sometimes useful starting point.

[115] Allen, supra note 58, at 272.

be used) to relate the conjunctive probability to the probability of individual elements. For example, for a two element claim, the multiplicative property states:

$P(A \text{ and } B) = P(A) \times P(B|A),$

while for a three element claim,

$P(X \text{ and } Y \text{ and } Z) = P(X) \times P(Y|X) \times P(Z|X \text{ and } Y).$

Using such properties, Allen and Leiter reiterate the claim that Criterion 1, together with its corresponding extension to more than two elements, produces arbitrary results based on the fortuity of the number of elements into which the cause of action is broken down:

> Take the example of theft and murder. Theft has considerably more elements than murder. Thus to convict for theft requires on average that intent to steal [for example, element X above] be established to a higher probability than intent to kill [for example, element A above] for a murder conviction.[116]

However, Professor Richard D. Friedman's earlier reply to the same argument demonstrates that the apparent paradox is a mirage if one keeps in view the substance of the claim rather than the elements by reference to which its legal sufficiency is assessed:

> [A]ssuming a claim is not altered substantively, dividing the claim into more elements will in fact raise the average probability that a fact-finder would assign to each element. Because the redivision of the claim has not altered its substance, the fact-finder's assessment of the probability of the truth of the entire claim cannot have changed; it follows that the average of the probabilities that the fact-finder assigns to each element must rise. Looked at another way, a corollary of the division of a claim into more elements without altering the substance of the claim is that there is less content in each of the elements. The average probability of the elements would therefore be ex-

---

[116] Allen & Leiter, supra note 1, at 1504–05.

pected to rise, especially given that the elements are not independent.[117]

To revert to Allen and Leiter's example, even if it is true that intent to steal must be shown to a higher probability than intent to kill, that does not mean that the prosecution's overall burden of persuasion is higher in a theft case than in a murder case.[118] All that one can say is that, under an expanded version of augmented Criterion 1, the division of a given cause into a greater number of elements means that there will be a larger number of subsidiary tests to be satisfied, each of which will be easier to satisfy.[119]

There are further advantages to the proposed synthesis. Distinguishing between decision rules and inferential rules of thumb allows the relative plausibility theory to accommodate other complications that are embarrassing to the theory as put forth by Allen and Leiter. For example, a serious argument can be made that "more likely than not" is not the best standard of proof for all civil cases, even simple two party cases, and indeed, it is not employed in all such cases. As the Supreme Court's due process cases demonstrate, it depends on the costs associated with false positives and false negatives.[120] Even for garden variety civil cases, the "more likely than not" rule is often dubious. While it makes sense in cases of strict, no-fault liability, it is not demanding enough for cases in which liability entails publicly labeling someone as having breached a serious moral obligation, such as those alleging fraud,

---

[117] Richard D. Friedman, Answering the Bayesioskeptical Challenge, 1 Int'l J. Evidence & Proof 276, 283 (1997). Indeed, this illustrates once more why Criterion 1 is superior to Criterion 3. As Friedman argues, if one uses an element-by-element decision criterion like our Criterion 3: "[T]he more elements a claim is divided into, the easier it is for the plaintiff to satisfy the burden. . . . Thus, defining the standard of persuasion in terms of individual elements becomes incoherent." Id. at 280.

[118] If anything, one might expect that the jury would apply the same nominal "beyond reasonable doubt" standard in such criminal cases by recognizing that what might be reasonable doubt in a murder trial might not be reasonable doubt in a theft trial. See supra note 77.

[119] Due to random variance in the assessment of probability as to each element, the sheer multiplicity of tests might offset the increasing average probability for the truth of individual elements by presenting a greater chance of getting at least one determination on a particular element that is adverse to the party with the burden of persuasion. This second order effect does not significantly undermine Friedman's basic point.

[120] See sources cited supra note 78.

gross negligence, and perhaps even ordinary negligence—and the law often reflects this fact.[121] Allen and Leiter have not suggested an intelligible way for the relative plausibility theory to handle these kinds of adjustments and remain a decision rule.[122]

The expected utility theory can readily incorporate this kind of consideration because such cases involve a disutility from a false positive (for example, an erroneous statement that defendant breached a serious duty of care) that exceeds the disutility from a false negative (for example, an erroneous statement that an injury was accidental), producing a critical probability greater than 0.5.[123] Further, if an elevated standard of proof were viewed as necessary for only one of several elements of a cause of action, this situation can also be accommodated by employing Criterion 1 augmented by a requirement—which could easily be conveyed in a jury instruction—that the particular element must be established by the higher standard.[124] Once again, the comparative plausibility of the parties' stories is a useful place for the jury to begin assessment, but that theory cannot do all the necessary work.

In one respect, however, this synthesis does not appear to accomplish the epistemological goals of the relative plausibility

---

[121] See Dale A. Nance, Civility and the Burden of Proof, 17 Harv. J.L. & Pub. Pol'y 647, 659–72 (1994).

[122] But see Allen, supra note 61, at 413 (claiming that the relative plausibility theory can accommodate the intermediate standard "clear and convincing proof" and translating it as "a considerably more persuasive story than its opposition," but not explaining how much more is "considerably more" and acknowledging that this is "the most troublesome standard of proof for this theory").

[123] Allen and Leiter might object that once one moves away from the "0.5 rule" one cannot make a transformation analogous to that from Criterion 1 to Alternative Criterion 1, and thus one must require jurors to calculate an actual number. In most cases it is true that jurors do not, and perhaps cannot be expected to, "calculate probabilities" in deciding a case, but it is important to remember that the actual specification of a number, whether by calculation or intuition, is not a necessary feature of these criteria. As Professor Peter Donnelly has observed, all that is necessary to the application of such criteria is to decide whether one probability is greater than another, even when the "other" probability is quantified: "As in the quantitative sciences, the task of establishing that some quantity falls (say) above a particular value is easier, often enormously so, than an exact evaluation of the quantity." Peter Donnelly, Approximation, Comparison, and Bayesian Reasoning in Juridical Proof, 1 Int'l J. Evidence & Proof 304, 306 (1997).

[124] A common example of this kind of situation is the requirement that malice be proved with clear and convincing evidence in order to warrant punitive damages. See, e.g., Tuttle v. Raymond, 494 A.2d 1353, 1363 (Me. 1985).

theory, and for that reason, if no other, it may be unacceptable to Allen and Leiter. Recall that a principal virtue claimed for the theory is that it offers a way out of the bleak prospect of the jury being required to attend to all possible explanations, including an enormous range of low-probability stories, lest some accumulation of such stories affect the decision one way or the other. Insofar as the proposed synthesis requires attention to third stories, that prospect remains troublesome. The key in Allen and Leiter's discussion appears to be the suggestion that the jury limit its attention only to a much smaller set of "plausible" stories. Nevertheless, Allen and Leiter provide us no real help as to how such a restriction of attention is to be achieved.

In this regard, remember what Allen and Leiter have to say about criminal cases: "[I]n criminal cases the prosecution must provide a plausible account of guilt and show that there is no plausible account of innocence."[125] It is difficult to quarrel with this statement as a reading of the "beyond reasonable doubt" standard of criminal law. But questions quickly emerge. Most obviously, in what respect is this an instantiation of a *relative* plausibility theory? How does this rule follow from the idea that "legal factfinding involves a determination of the comparative plausibility of the parties' explanations offered at trial"?[126] There is simply nothing *comparative* about the test Allen and Leiter state for criminal cases, because "plausibility" here is necessarily assessed in the abstract, not as compared to some other story.[127] Allen and Leiter thus avoid the problem of the third story only at the cost of removing criminal law from the relative plausibility framework. And there is no clue here as to how, even in principle, one can determine how probable the defendant's story must be in order to be plausible or in what other way the jury is to decide whether a story is plausible.

---

[125] Allen & Leiter, supra note 1, at 1527–28.

[126] Id.

[127] Presumably this shift is necessitated, at least in part, by the narrower availability of discovery in criminal cases. Recall their argument: "In civil cases, given mutual discovery, the parties can be expected to search for and produce evidence of whatever stories they think can plausibly support their legal claims." Allen & Leiter, supra note 1, at 1531.

Admittedly, the criminal standard of proof is notoriously difficult to interpret. Nevertheless, if one construes the "beyond reasonable doubt" standard as a common-sense, non-quantitative way of expressing the result of a calculation under the expected utility theory, something like "greater than 95% probability," then *at least in principle* there is a standard by which to assess whether reasonable doubt exists: When combined with third stories that neither party tells, defendant's explanation should identify a 5% probability of innocence. Once again, in making these assessments, whether quantified or not, a comparison of the degree of plausibility of the prosecution story and the defense story is a very useful starting point. The jury is well on its way to a verdict if there is a consensus that the prosecution's story is twenty times more likely than the defendant's. Nonetheless, the possibility of the third story means that, even in such a case, the jury's work is not quite finished.[128]

We shall never have a complete account of factfinding until we know much more about the question of what makes a story plausible, or rather plausible enough to be considered with respect to the task at hand, and Allen and Leiter should be praised for insisting that we give attention to this issue. At the same time, this missing piece in the theory of factfinding need not cause us to reject the synthesis suggested here. We certainly know, by introspection if nothing else, that people required to explain some event or evidence are able to restrict their attention to the most pertinent and promising potential accounts—at least, they *believe* they are doing that—and we need not know exactly how this is done in order to formulate decision criteria in probabilistic terms. Perhaps some rough and ready rules of thumb will emerge to the effect that, for the practical purposes of law, only the three or five most plausible stories need be considered, and perhaps we can adjudicate using practical reason without having to explain just how the jury identifies these most plausible stories starting from the two that the

---

[128] An instructive example in the criminal context is that of the innocent defendant who is not willing to incriminate the person he knows to be guilty but who does not want to go to jail himself. The prosecution's story might well be twenty times more likely than the best story such a defendant is willing to produce, even though the jury can identify a quite plausible story (incriminating a third person) that is not advocated by either side.

parties present. Any such rule of thumb, however, would remain in the domain of inferential practice. It *need* not be a part of the formal decision criterion, if only because we can rely on the common sense of the jury in implementing it, and it *should* not be a part of such a criterion because we would have enormous difficulty articulating such an idea in a way that would not unnecessarily restrict the free play of inferential thinking that is critical to effective decisionmaking.

III. BAYESIAN ANALYSIS OF PROBATIVE VALUE

A. Bayes' Theorem Applied to Trials

Another branch of decision theory relates not to the burden of persuasion criterion but rather to the analysis of the probative value of evidence. Bayes' Theorem, derived from the axioms of probability, relates the assessment of the probability of an event given certain information to the assessment of the probability of the same event without such information.[129] Let $P(G)$ represent the

---

[129] The referenced axioms are the so-called Kolmogorov axioms, formal requirements of a system of probability measurement that prevent self-defeating sets of beliefs about uncertainty. See, e.g., Brian Skyrms, Choice and Chance: An Introduction to Inductive Logic 168–98 (2d ed. 1975). Not all scholars believe that this system of probability is the correct one to use in thinking about problems of inductive inference, especially problems like inference in legal trials. The best known such challenge is that of Jonathan Cohen. See Cohen, supra note 67. The argument presented by Cohen and those who endorse his view tends to focus on the problem of evidential incompleteness. The idea is that when we have very little evidence about the occurrence of an event, A, then we are not warranted in giving a high level of probability to either A or not-A. Id. at 33–47, 74–86, 171–81, 219–24, 270–73. This argument seems to conflict with one of the fundamental axioms or theorems of mathematical probability, namely that, for any event A, $P(A) + P(\text{not-}A) = 1$, which requires that when $P(A)$ is low (close to 0), $P(\text{not-}A)$ must be high (close to 1). See, e.g., Alex Stein, Judicial Fact-finding and the Bayesian Method: The Case for Deeper Skepticism About Their Combination, 1 Int'l J. Evidence & Proof 25, 28–33, 41–43 (1997). The matter is too complicated to address here in detail, but my present view is that this critique of the use of mathematical probabilities, and a fortiori of Bayes' Theorem, in the legal context fails to distinguish between (1) the issue of whether it is justifiable to make an assessment of probability for a given decision task on the evidence then available and (2) the question of what probability to assign if assignment is to be made. See Nance, supra note 46, at 625 (drawing the indicated distinction, arguing that it is the trial judge's duty to assure that the evidence is sufficiently complete to warrant submission of the case to the trier of fact for decision in accordance with an assessment of probability, and recommending how the trial judge should discharge that duty as part of the burden of production).

probability that the facts would be such that defendant is guilty (or liable), considered without regard to a particular piece of evidence, E, and let $P(G|E)$ represent the probability of guilt taking E into consideration. Further, let $P(E|G)$ represent the conditional probability, assessed without regard to the fact that it is received, that the evidence E *would be* received given that the defendant *is* guilty in fact, and $P(E|\text{not-}G)$ represent the conditional probability, assessed without regard to the fact that it is received, that the evidence E *would be* received given that the defendant is not guilty in fact. Then under Bayes' Theorem:

$$P(G|E) = \{P(G) \times P(E|G)\} \div \{[P(G) \times P(E|G)] + [P(\text{not-}G) \times P(E|\text{not-}G)]\}$$

A mathematically equivalent, but more transparent and often more useful version of the rule is:

$$O(G|E) = O(G) \times L_G(E)$$

or in words,

    posterior odds = prior odds × likelihood ratio

where $O(G)$ is the "prior odds" of guilt (or liability), which is the ratio of $P(G)$ to $P(\text{not-}G)$, $O(G|E)$ is the "posterior odds" of such guilt (liability), or the ratio of $P(G|E)$ to $P(\text{not-}G|E)$, and $L_G(E)$ is the "likelihood ratio" for evidence E relative to the hypothesis G, or the ratio of $P(E|G)$ to $P(E|\text{not-}G)$. The likelihood ratio represents the relative compatibility of the evidence E with the two competing hypotheses, G and not-G, and constitutes a measure of the probative value of E relative to these hypotheses.[130]

---

[130] A more detailed statement of the theorem, explicitly taking into account background information affecting the conditional probabilities, can be found in C.G.G. Aitken, Statistics and the Evaluation of Evidence for Forensic Scientists § 2.5.1, at 46–50 (1995). A similar treatment may be found in Ian W. Evett & Bruce S. Weir, Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists 22–29 (1998).

In theory, one can iterate this equation for each piece of evidence considered, so that the entire case could be represented by an equation of the following sort:

$$O(G|E) = O(G) \times L_G(E_1) \times L_G(E_2|E_1) \times \ldots \times L_G(E_n|E_1 \ldots E_{n-1})$$

where E now represents the total package of n items of evidence received, E_1 through E_n, where $L_G(E_i|E_j)$ is defined in the obvious way, as the ratio of $P(E_i|G$ and $E_j)$ to $P(E_i|\text{not-}G$ and $E_j)$, and where $O(G)$ now represents the odds in favor of guilt (liability) without any evidence, or at least any evidence on contested facts.[131] This equation does not assume that the E_i are presented to the decisionmaker in any particular temporal order, nor does it assume that there is a *unique* decomposition of the total evidence into n pieces. In the general case, however, the likelihood ratio associated with a particular piece of evidence does depend on what *other* evidence has already been taken into account. That is, it depends on the sequence in which evidence is *taken into account* (rather than the sequence in which the evidence is *presented*). For example, if one item of evidence is that the perpetrator, like the defendant, was tall, the value of the likelihood ratio for that evidence can depend on whether this evidence is considered first, as E_1, or considered second, as E_2. If considered second, after the consideration of evidence that the perpetrator (like the defendant) was male, the likelihood ratio for "perpetrator-was-tall" is smaller than it would be if the reasoner has not already taken into account "perpetrator-was-male." At the same time, the likelihood ratio for the *combination* of the two items of evidence (such as that the perpe-

---

[131] Characterizing $O(G)$ presents some interesting problems, especially in view of the "presumption of innocence" in criminal cases. The obvious problem is that some consider the natural meaning of "presumption of innocence" to be a prior probability of zero. In that case, $O(G) = 0$, and no evidence, no matter how probative, can raise the prior probability, resulting in a posterior odds and posterior probability of zero as well. A number of strategies have been suggested for dealing with this problem. Compare, e.g., Posner, supra note 12, at 1514 (stating that an unbiased juror should start with one to one odds of guilt in a criminal case or liability in a civil case), with Richard D. Friedman, A Presumption of Innocence, Not of Even Odds, 52 Stan. L. Rev. 873 (2000) (criticizing Posner's account of starting odds and employing instead the factfinder's hypothetical pre-indictment, pre-arrest assessment of the odds of the defendant's guilt).

trator, like the defendant, was a tall male) is not affected by the order in which its two pieces are taken into account. Moreover, if two items of evidence are stochastically independent, then the likelihood ratio of each is unaffected by the sequence in which they are taken into account, ceteris paribus. Thus, if E_i through E_n are mutually independent, then the iterative equation presented above simplifies[132] to:

$$O(G|E) = O(G) \times L_G(E_1) \times L_G(E_2) \times \ldots \times L_G(E_n)$$

Whether or not the E_i are mutually independent, in theory the final $O(G|E)$ should not depend on how the individuation occurs or in what order the evidence is presented or considered, although in practice such differences might matter psychologically.[133]

Much has been written in the last twenty years or so about the usefulness of this kind of representation of the assessment of evidence at trial.[134] On the one hand, even its staunchest proponents do not claim that the equation represents a complete account of the process of assessing $P(G|E)$, whether as a descriptive or prescriptive matter. In particular, it is well understood that neither Bayes' Theorem, nor any of the many other theorems derived from the axioms of probability, determine by themselves the value of $P(G)$ or the likelihood ratios, and consequently, they cannot determine $P(G|E)$. Rather, those theorems provide normative consistency constraints on the assessment of such magnitudes.[135] On the other hand, even its staunchest opponents do not categorically deny that Bayes' Theorem might be useful in at least some ways in thinking about the assessment of evidence at trial, though they con-

---

[132] See Aitken, supra note 130, § 5.1.3, at 110–16.

[133] The stated psychological qualification, if true, merely implies what no Bayesian denies: that Bayes' Theorem is not a complete account of the actual inference processes that take place at trial.

[134] A symposium addressing these issues, organized by Professor Allen, appeared as a special issue of a (then) new journal. Symposium, 1 Int'l J. Evidence & Proof 253 (1997). An impressive book length treatment is David A. Schum, The Evidential Foundations of Probabilistic Reasoning (1994).

[135] See, e.g., Richard D. Friedman, Towards a (Bayesian) Convergence?, 1 Int'l J. Evidence & Proof 348, 349–51 (1997).

tinue to express strong skepticism.[136] Rather, they see the consistency constraints as simply too weak to be of any serious help in modeling, descriptively or prescriptively, the actual inferences of factfinders.[137]

Given this state of the debate, one might argue that there is little left to say besides the discussion of particular analyses proposed by those who consider Bayes' Theorem useful. In one respect, however, Allen and Leiter might well be expected to see some general value in attending to likelihood ratios. Bayesian analysis is just as compatible with the relative plausibility model advocated by Allen and Leiter as it is with an expected utility theory of the burden of persuasion. Indeed, many Bayesian analysts consider it very useful, as a practical matter, to specify a competing hypothesis (story) in place of the more abstract competing hypothesis "not guilty" or "not liable" (that is, not-G). It is often easier to assess the likelihood ratio for particular evidence if the defendant articulates a theory of the case and one can restrict attention to that particular story instantiating "not guilty."[138] Despite this seeming affinity, Allen and Leiter continue to argue generally against the use of Bayes' Theorem in analyzing evidence at trial.[139] Accordingly, some brief responses of comparable generality are in order.

B. Some Responses to Allen and Leiter's General Arguments

In their paper, Allen and Leiter briefly reiterate several claims about the use of Bayes' Theorem in the juridical context. First, they argue that such Bayesianism founders on the problem of

---

[136] This is made clear in Allen and Leiter's discussion of Posner's use of Bayesian analysis. See Allen & Leiter, supra note 1, at 1519–20.

[137] See, e.g., Allen, supra note 58. This point is made particularly well, and succinctly, in Craig R. Callen, Computation and Juridical Proof, 1 Int'l J. Evidence & Proof 296 (1997). There is an interesting parallel here with the difference between those who see formal justice, the idea that the law should treat like cases alike, as an empty concept, and those who attribute significant force to it in shaping the law. Compare Peter Westen, The Empty Idea of Equality, 95 Harv. L. Rev. 537 (1982) (treating formal justice as an empty concept), with Ronald Dworkin, Law's Empire (1986) (attributing significant force to formal justice).

[138] See Bernard Robertson & G.A. Vignaux, Interpreting Evidence: Evaluating Forensic Science in the Courtroom 33–50 (1995) (discussing the problem of identifying the "alternative hypothesis").

[139] Allen & Leiter, supra note 1, at 1507–10.

computational complexity—that to actually carry out the computations implicated by Bayes' Theorem for any ordinary litigated case would be impossibly complex.[140] This frequently stated objection has always struck me as oddly misdirected. If the point is simply to counsel against a general practice of having the jury try to analyze the entire mass of evidence in a case by explicitly using the theorem and (presumably) a calculator to obtain a precise posterior probability, then the point is correct but not particularly interesting. Such a procedure is obviously impractical, and no Bayesian enthusiast of whom I am aware suggests that we implement it.[141] It does not follow from that point, however, that Bayes' Theorem is not useful, either descriptively or prescriptively, in modeling jury decisionmaking.

Consider the descriptive issue. Is it true that a process with a measurable output cannot be usefully described by a formal model unless there exists within the modeled process an intelligence capable of performing the calculations presented in the model? Clearly not. Physical systems are often modeled in science with theories requiring considerable computation even when the physical system does not, to our knowledge, contain any intelligence at all, much less an intelligence capable of performing the necessary calculations. Even if a human intelligence is present and operating in the modeled process, the model, to be useful, does not need to be one that the modeled intelligence can or does consciously employ. Think of the physics of riding a bicycle or throwing a baseball.[142] Indeed, the formal representation of the heuristics used in practical decisionmaking can be, and often is, quite complex.[143] That does not mean that we cannot learn about the decisionmaking process by developing such complex representations, even if many

---

[140] Id. at 1507.

[141] Of course, this does not rule out as necessarily impractical the use of Bayes' Theorem in court, in a limited class of cases, to illustrate the probative value of certain evidence. This is discussed infra Section III.D.

[142] Cf. Friedman, supra note 117, at 289 ("[W]hen thinking well—and with the aid of whatever simplifications and heuristics may be necessary—factfinders reach results that are roughly consistent with those they would reach if they were to apply probability theory rigorously. They do not have to think about the theory consciously, just as an athlete does not have to think about the laws of physics in determining where a ball hurtling through the air is likely to land on the ground.").

[143] See generally Gigerenzer et al., supra note 57 (examining various heuristics).

of those who employ the heuristics are incapable of understanding the corresponding representation.

Turning to the prescriptive issue, the objection might be made that such models, at least to the extent that they are not and cannot be employed by the modeled intelligence, are radically separated from that intelligence, so that the model is useful only "from the outside," for an observer of the process. Any appraisal done by such an observer must be merely passive: It cannot, the objection would go, tell the modeled intelligence anything of practical utility that might improve the results of the activity being modeled. But this is simply not so, as the physics of bicycle riding or baseball throwing could illustrate. An analyst could use tools that the intelligence being modeled does not or could not use, at least not explicitly, to reach recommendations about how the modeled activity can be improved. Similarly, the success or failure of a bridge can be analyzed using a particular model whether or not the designers or builders of the bridge were even aware of the model, as when an ancient design is studied using modern theories, and the results in turn can be used to improve modern bridge building. Moreover, even a limited degree of understanding of the model by a modeled intelligence can sometimes be of use in the actual decision process even if precise computations are infeasible. An engineer, for example, can make simplifying assumptions that make practical the computation of parameters for the construction of a bridge. Simply being familiar with the basic logic or structure of the formalizations of bridge construction theory can provide important insights for the designer or builder without requiring any serious calculations at all.[144] The variety of potentially helpful information feedback

---

[144] In a similar vein, Richard Friedman illustrates various useful general propositions about evidence, derived from Bayes' Theorem, the use of which requires no explicit calculation:

> 1. All other things being equal, the more probable a proposition appears without consideration of a given body of evidence, the more probable it will appear upon consideration of that evidence.
> 2. All other things being equal, the more probable it appears that a given body of evidence would arise given the truth of a proposition, the more probable the proposition will appear given the body of evidence.
> 3. All other things being equal, the less probable it appears that a given body of evidence would arise given the falsity of a proposition, the more probable the proposition will appear given the body of evidence.

Friedman, supra note 135, at 350.

mechanisms, between modeled intelligence and model builders, is considerable.

It is thus odd how intensely skeptical the anti-Bayesians are about the ability to identify useful formalisms for the juridical world. It is as if someone had said to Newton:

> How foolish you are to think that you can gain insight into the motions of physical objects with an equation like $f = ma$. It may be theoretically interesting, but you'll never be able to apply it in the real world, with all its enormous complexities, with countless forces acting on any object and an inability to capture precise initial conditions at some fixed point in time. What is more, nearly all the objects you think are governed by this equation cannot possibly do any of the required calculations, so how can they know how fast to accelerate!

As this *reductio* suggests, in the end one must look to the results that are generated by models to see if they are useful rather than ruling out the enterprise on a priori considerations such as computational complexity. Maybe, however, the argument from computational complexity is really aimed at a different point. To assess this possibility, I turn to the other general arguments mentioned by Allen and Leiter.

Allen and Leiter's second argument is that trials do not proceed in the manner contemplated by Bayesian revisions of probability. That is, trials do not proceed by the jury assigning a prior probability, introducing some new evidence, reassessing the probability of guilt, introducing further new evidence, reassessing the probability of guilt, and so on. Rather, the jury assesses the likelihood of guilt, or the parties' stories, only after all the evidence is in; at that point there is no "prior" probability, there is only "posterior" probability. [145] Like many anti-Bayesian arguments, this one has, at least to some extent, created a straw man to attack. Despite the somewhat misleading conventional terminology referring to "prior" and "posterior" probabilities, Bayes' Theorem does not say anything about *when* the assessments of probability are made, nor does it assume the evidence is assessed in the order in which it is presented to the trier of fact. One can apply Bayes' Theorem *after* all evidence has been introduced and considered. It is simply a way of decomposing

---

[145] Allen & Leiter, supra note 1, at 1507–08.

the evidence considered, and as noted above, there is no unique decomposition.[146]

This might not seem to be responsive to the point Allen and Leiter are making, as it merely shifts the problem of adjusting for complex dependencies to a different sequence of evidence. Allen and Leiter might still reject as unrealistic the idea of sequential evaluation, whatever sequence is selected. But even if that sequential evaluation were required to replicate in some close way the actual psychological states of the jurors in reaching a decision—an unnecessary restriction—allowing for flexibility to rearrange the order in which evidence is assimilated presents intriguing possibilities more compatible with a holistic approach. For example, strategies may exist that allow jurors to reduce the complexity of the task by grouping particular pieces of evidence in simplifying ways.[147] Such strategies may be more or less conscious and could utilize evolved heuristics for processing complex information. All this remains to be explored, and it is too constraining of inquiry to reject such possibilities a priori on the ground that equations like those presented above do not *look* like what jurors are doing when they think about a case.

In any event, nothing about Bayesian formalisms requires that *every* item of evidence be evaluated explicitly in Bayesian terms. In particular, Bayes' Theorem can be used simply to illustrate how the prescriptive effect of a single item (or collection of items) of evidence can be extracted from that of all the rest. To employ Bayes' Theorem utilizing a "prior probability" relative to a particular item

---

[146] Allen and Leiter comment in a footnote that their second argument is more forceful when applied to the problem of "discovery" than to the problem of "justification," but they claim that "the task at trial is more analogous to discovery than justification." Allen & Leiter, supra note 1, at 1508 n.51. To the contrary, the presentation of evidence at trial is decidedly justificatory, especially in an adversarial system in which each side tries to justify inferences favorable to that side. Moreover, jurors at trial must not only form (that is, "discover") beliefs about the likelihood of various events, but they must as a practical matter be prepared to offer justifications to their fellow jurors for those beliefs during deliberations—deliberations that can change an assessment that appears to be irrational. At each of these stages of discovery and justification, Bayesian decomposition of the evidence can potentially play a part.

[147] For example, two items of evidence that have the effect of canceling each other out might be grouped together and then ignored. I am indebted to Bernard Robertson for suggesting this line of thought.

of evidence, it suffices that the person employing it be able to articulate and answer the question, "What would I believe if I did not have that item of evidence but I did have all the rest?"[148] This allows one to assess the likelihood ratio for that item (or collection of items) as if it were the last piece of evidence taken into account. The likelihood ratio measures the marginal contribution of that piece of evidence to the entire evidence package.[149]

The third argument recited by Allen and Leiter emphasizes a point conceded by Bayesians, that Bayes' Theorem provides only a set of consistency constraints on probability assessments and so cannot be enough, by itself, to generate the posterior probability of interest. Defending Bayesian analysis, Friedman has commented on the limited significance of this point as follows: "[T]his consideration does not undermine the value of the theory—any more than the laws of physics are rendered useless because they do not reveal the mass of a given object but only indicate what happens in prescribed conditions to an object of a given mass."[150] But because what is "constrained" by such probability theory is a set of *subjective* probabilities, Allen and Leiter remain unconvinced by such replies:

> [I]ndividuals can begin from radically different perspectives, and each, in Bayesian terms, will be operating equally rationally. Bayes' Theorem provides no method of adjudicating such differences and thus cannot offer useful guidance for factfinders. In other contexts, such as science, these differences may be marginalized by convergence theorems that demonstrate that over time and with enough new evidence the divergent initial

---

[148] Of course, one must be alert to the fact that framing the question this way may introduce the kind of distortion associated with hindsight bias. See generally Scott A. Hawkins & Reid Hastie, Hindsight: Biased Judgments of Past Events After the Outcomes Are Known, 107 Psychol. Bull. 311 (1990) (reviewing empirical evidence of this phenomenon).

[149] It is conceivable, of course, that isolating a single item of evidence for analysis in this fashion might somehow be incompatible with a more holistic evaluation of all the evidence. For example, the former might fail to capture some aspect of the evidence that is useful under the heuristics by which the latter is performed. But merely being able to conceive of such a possibility is not enough to rule out of bounds the type of isolating analysis discussed in the text; that would take some greater showing that such analysis would yield results, such as the exclusion of evidence, that would distort the jurors' understanding in a way that is, on balance, veritistically counterproductive.

[150] Friedman, supra note 135, at 350.

starting points will wash out and the result will converge on the truth. There is nothing even remotely analogous to this in the condition of trials.[151]

Is this true? At the individual level, we do have the beginnings of some convergence results, results that tie subjective assessments to accurate results. Alvin Goldman provides the following commentary in his recent book, in a section strikingly entitled, "A Veritistic Rationale for Bayesian Inference":

> Good inference procedures alone do not guarantee veritistically good outputs; one also needs good factual inputs. ... No deductive method can pledge to a reasoner that its use will guarantee true conclusions. Only a more modest claim can be made: true conclusions will follow if the reasoner's *premises* are true. ...
>
> If this is the strongest claim that can be made on behalf of deductive inference, more can hardly be expected from inductive or probabilistic inference. One should be pleased to find any analogous property in the latter domains. Precisely such an analogous property is what Bayesian inference can be shown to possess, under [certain] assumptions .... [W]hat I shall show (roughly) is that when a reasoner starts with *accurate likelihoods* [that is, accurate likelihood ratios] (analogous to true premises), it is objectively probable that Bayesian inference will increase his degree of knowledge (truth possession) of the target proposition.[152]

Goldman's main point is that this tendency toward "truth possession"—a variable property that essentially means assigning high subjective probabilities to propositions that are true and low subjective probabilities to propositions that are false—does not depend on the magnitude of the individual's prior (subjective) probability about the target proposition.[153] This result, which

---

[151] Allen & Leiter, supra note 1, at 1508.

[152] Goldman, supra note 43, at 115–16. Goldman subsequently states his results in the form of two theorems. Id. at 121–22.

[153] Goldman further summarizes his result:

> We do not assume, however, that the reasoner begins with a particularly good or bad estimate of the truth or falsity of X [the target proposition]. Indeed, our analysis will show that wherever she starts—whatever her *prior* probabilities for X and for NOT-X—application of Bayes' Theorem to derive a *posterior* probability in light of the witness's testimony leads to an (objectively) expected

Goldman illustrates after the quoted passage, obviously depends on the assumption that subjective juror likelihood ratios are objectively accurate. In that respect, the demonstration is incomplete.[154] But it is just the sort of thing Allen and Leiter appear to demand from Bayesians. To carry the matter further and address the accuracy of a juror's likelihood ratios, one would need to look to the effects of experience (including the effects of natural selection and cultural evolution) and, when expertise is helpful, to the effects of expert assistance. In terms of "adjudicating" differences among individuals, including jurors, one would also need to look to the effects of interpersonal dialogue, such as jury deliberation.

These are just the beginnings, as I said, but this discussion suggests that the situation for Bayesianism, though certainly complex, is not nearly so hopeless as Allen and Leiter suggest—at least as long as one does not demand too much of the rules of inductive inference. As Goldman's statement serves to emphasize, we do not reject the juridical use of rules of deductive logic, either as bases for argument by counsel or as tools of persuasion by jurors in deliberations, even though merely conforming to such rules does not guarantee accuracy of results. The same is true of Bayes' Theorem and the general results of probability theory of which Bayes' Theorem is a part. Carefully employed, they are tools with potential value in the very process that Allen and Leiter recommend, namely, the "painstaking attention to and examination of the evidence and its logical and empirical implications."[155]

I hope that what I have said, together with similar comments by others, will help to defuse the *general* attack on Bayesian thinking as a tool of probably limited but not insignificant—and still underexplored—utility. In the following two Sections, I turn to the

---

increase in her truth possession. (More precisely, this holds so long as her prior probabilities are neither 0 nor 1.0, and the likelihood ratio is not identical to 1.0.)
Id. at 118.
[154] Most importantly, the meanings of terms like "objective probability" and "objectively accurate likelihood ratios" require elaboration. See id. at 117. Goldman has undertaken this elaboration in an intriguing recent paper. See Alvin I. Goldman, Quasi-Objective Bayesianism and Legal Evidence (2001) (unpublished manuscript, on file with the Virginia Law Review Association).
[155] Allen & Leiter, supra note 1, at 1509. Obviously, some particular uses of Bayesian analysis might present the "allure of the false hope" Allen and Leiter fear. Id. That can only be assessed in the context of the particular suggested uses.

residual undertaking, suggesting specific applications of Bayesian analysis that show value in the juridical context. I offer these as the testimony of someone who by no means considers Bayesian analysis to be the *only* useful tool. Conversely, there are other possible sources of value in the Bayesian approach, so I do not claim to exhaust the possibilities. Instead, I focus on the sources of value that I have encountered in my work as a teacher and as a scholar.

C. Bayes' Theorem as an Analytical Tool for Lawyers

As already noted, critics of Bayesian analysis, including Allen and Leiter, do not deny that such analysis might serve as a useful heuristic for some purposes other than modeling the actual inference process at trial. Such purposes might include informing a lawyer's analysis of relevance and probative value in the context of decisions about conducting discovery, arguing to the judge about admissibility, arguing to the jury about weight, or arguing before a rulemaking authority about the rules of admissibility. A number of scholars have thought this kind of analysis is useful enough to justify the effort to undertake it.[156] I will not reexamine the arguments made by such individuals; their work speaks for itself. Rather, in this short section, I want to indicate what I have learned about the *pedagogical* utility of Bayesian analysis.

When students in a basic evidence course are asked to assess the relevance or probative value of some item of evidence offered by the prosecution in a criminal case, the all too common classroom response is to say that the evidence is (or is not) relevant because the evidence does (or does not) make it more likely than not that the defendant committed the alleged crime. When pressed on the matter, students will see that this verbalization is imprecise since relevance does not mean that the evidence in question causes the ultimate fact of interest to reach any particular level of probability; it only means that the probability is affected.[157] They will correct themselves by stating that the evidence is (or is not) relevant be-

---

[156] See, e.g., V.C. Ball, The Moment of Truth: Probability Theory and Standards of Proof, 14 Vand. L. Rev. 807 (1961); Richard Friedman, Character Impeachment Evidence: Psycho–Bayesian [!?] Analysis and a Proposed Overhaul, 38 UCLA L. Rev. 637 (1991); Kaplan, supra note 46; Richard O. Lempert, Modeling Relevance, 75 Mich. L. Rev. 1021 (1977).

[157] See Fed. R. Evid. 401.

cause it does (or does not) make the defendant's guilt more likely than it would be without the evidence. So far, so good. But when students are pressed to explain *why* that is so, they often have little or no ability to support their conclusion except to say that it seems that way to them.

For example, consider evidence by a witness that defendant was observed running from the scene of a crime as the police approached. Students readily assess this evidence as relevant but often have difficulty articulating just why. When pressed, they will sometimes say that running from the scene shows the defendant's desire not to be caught for having committed the crime. When pressed to explain whether it might show something else, they will begin to give explicit attention to alternative hypotheses. Was the defendant scared of the police for reasons having nothing to do with the crime at issue? Did he have reason to be? Might he have been afraid of retaliation if he were forced to give evidence against the true perpetrator? And so on. Students then come to see that the probative value is dependent on the relative strength of the consciousness-of-guilt explanation and the other possibilities.

Without ever seeing a formula on the blackboard, students have thus internalized a more explicit understanding of the core concept of Bayesian thinking. They are examining the relative plausibility of competing explanations of the particular evidence, just as Bayes' Theorem suggests they should.[158] Having been sensitized to the need for such inquiry, students often can employ it in other contexts. For example, when subsequently presented with a personal injury case in which there is evidence that the plaintiff attempted to suborn perjury from a witness to support his claim, students are much quicker to look for explanations of such an attempt other than the obvious one. In the case that I use, it is not difficult to find information supporting such an alternative theory of the plaintiff's behavior, once one thinks to look for it.[159] Such an inquiry provides

---

[158] I have found that obtaining such internalization before presenting the formula (or even without ever presenting the formula) decreases student resistance to the idea presented, just as illustrating Coase's Theorem without ever stating it improves student understanding by decreasing resistance.

[159] The case is *McQueeney v. Wilmington Trust Co.*, 779 F.2d 916 (3d Cir. 1985). It presents the possibility that the plaintiff, while in the right on the merits, was moved to subornation because he worried that his legitimate claim might be frustrated in the courts by the defendant's unscrupulous delaying tactics. See id. at 919 n.2.

the student with a basis for arguing that the evidence of suborna-
tion is, if not irrelevant, at least less probative than might at first
glance appear, making more plausible an admissibility argument
based on the risk of unfair prejudice or an argument to the trier of
fact that the evidence should not be given much weight.

Obviously, this framework does not provide complete answers to
the questions that must be addressed in assessing probative value.
Knowing that one should compare $P(E|G)$ with $P(E|not\text{-}G)$ does
not answer the question of what the magnitudes of these probabili-
ties are or even what their ratio is. Once again, Bayes' Theorem is
not a complete guide to the justification of conclusions about rele-
vance or weight, but it is a useful starting point. Moreover, as
already noted, the Bayesian comparison is one that should be at-
tractive to Allen and Leiter, in light of their emphasis on the
relative plausibility theory. A focus on the likelihood ratio invites
just the kind of comparative assessment that the relative plausibil-
ity theory encourages, now directed at the question of the
probative value of a particular piece of evidence.

When one turns to scientific evidence, this kind of analysis be-
comes even more important. I have been teaching courses on
scientific evidence for the last three years, and I have found that
once students are exposed to Bayesian analysis, they are better
able to spot certain issues relating to the assessment of the proba-
tive value of such evidence. To take the obvious case, if an expert
offers to testify that a biological "mark" found at the scene of a
crime (a blood type, for example) "matches" the corresponding
characteristic of the accused, students familiar with Bayesian think-
ing naturally pose the question of what circumstances or events,
consistent with innocence, would explain the report of a match,
and then inquire how likely such circumstances or events are as
compared to the report of a match for an accused who is guilty.
Less obviously, if an expert offers to testify that a defendant or an
alleged victim displays behavioral features "characteristic" of a
battered person syndrome, familiarity with Bayesian analysis
prompts attention to the question of whether those features are
also present in those who do not fit the syndrome and, if so,
whether the expert can speak meaningfully to the relative fre-
quency with which the feature appears among those who do and

those who do not fall within the syndrome category—and if not, what significance the court should attach to that fact.

I will provide no further illustrations, because few if any Bayesioskeptics, as they are often called, deny the potential usefulness of thinking in these terms.[160] If they have any objection to this kind of use, it would seem to be that students and lawyers can appreciate the importance of asking the questions suggested above without ever attending to the Bayesian formalisms.[161] To that point, I would reply pragmatically: My experience has been that students do not tend to identify such questions *as readily* without being sensitized, at least implicitly, to the Bayesian framework.[162] The fact that there might be other ways to elicit comparable insights from students as readily does not negate the utility of this one. Of course, I concede that this is anecdotal evidence on pedagogical value put forth without the benefit of any systematic learning experiment or detailed knowledge of the relevant experience of other evidence teachers in this regard.

D. *Bayes' Theorem Used to Assess or Assist the Jury's Evaluation of Evidence*

There is general recognition that a Bayesian analysis might, at least in theory, be useful in the context of scientific evidence with an explicitly statistical component. For example, it was suggested many years ago that Bayes' Theorem might be used to convey the significance of the "random match probability" for forensic identification evidence used in criminal trials.[163] In a case in which a defendant matches a mark (for example, DNA profile or blood type) found at the scene of the crime, the random match probability is simply the chance that, though innocent, the defendant

---

[160] Indeed, one may plausibly infer that Allen considers the Bayesian paradigm to have significant pedagogical value since he includes a substantial discussion of it in his excellent casebook for the basic evidence course. See Ronald J. Allen et al., Evidence: Text, Cases, and Problems 191–97 (2d ed. 1997).

[161] This seems to be one of the arguments made by Professor Craig Callen. See Callen, supra note 137, at 298.

[162] As noted above, "implicit" sensitization may be more effective than "explicit," if by "explicit" one means actually writing a formula on the blackboard. See supra note 158.

[163] Michael O. Finkelstein & William B. Fairley, A Bayesian Approach to Identification Evidence, 83 Harv. L. Rev. 489, 502 (1970).

would—by mere coincidence—match that mark. This, in turn, is estimated by reference to the frequency of that mark in an appropriate suspect population.[164] This random match probability is a *part* of what contributes to the denominator of the likelihood ratio for the testimony reporting a match. Put another way, it is the denominator of the likelihood ratio for the proposition that there is a match relative to the hypothesis that the defendant is the source of the mark at the crime scene.[165]

The idea of Bayesian assistance is that jurors might be told the relative likelihood of getting a match under the hypothesis that the defendant is the source as compared to that of getting a match under the hypothesis that someone else (unknown and therefore statistically random as to the mark) is the source. For a random match probability of 0.04 (meaning that one out of twenty-five people on average share that mark), as an example, the expert would report to the jury that it is twenty-five times more likely that there would be a match if the defendant is the source than if some other unknown person is the source. (This presentation format will be called the "likelihood ratio format" in what follows.) Additionally, the same (or another) expert might illustrate for the jury the effect this likelihood ratio should have on various prior probabilities that the defendant is the source of the mark. That is, the jury might be shown a chart like the following, mapping prior probabilities (expressed as a percentage) to posterior probabilities for a likelihood ratio of twenty-five:

---

[164] Ian Evett and Bruce Weir make admirably clear the complexity of the attribution of a random match probability, without using that terminology. Evett & Weir, supra note 130, at 22–28.

[165] The numerator for this likelihood ratio is just 1, because it is (essentially) certain that defendant will match the mark at the scene if in fact it came from him. When taken together with the probability of a false positive laboratory error for an innocent defendant, the random match probability generates the denominator of the (different) likelihood ratio for the proposition that the forensic scientist *found* a match relative to the hypothesis that the defendant is the source of the mark at the crime scene. To obtain the denominator of the (still different) likelihood ratio for the evidence, which is a testimonial or documentary *report* of a match relative to the hypothesis of defendant's *guilt*, one must further take into account the chance of the expert witness lying, the chance of someone planting the mark at the scene to frame the defendant, the chance of there being an innocent explanation of the fact that the defendant *is* the source of the mark at the scene, and so forth. See Jonathan J. Koehler et al., The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?, 35 Jurimetrics J. 201, 203–05 (1995).

Prior Probability	→	Posterior Probability
0%	→	0%
5%	→	57%
10%	→	74%
15%	→	82%
20%	→	86%
25%	→	89%
30%	→	91%
35%	→	93%
40%	→	94%
45%	→	95%
50%	→	96%
55%	→	96.8%
60%	→	97.4%
65%	→	97.9%
70%	→	98.3%
75%	→	98.7%
80%	→	99.0%
85%	→	99.3%
90%	→	99.6%
95%	→	99.8%
100%	→	100%

(This presentation method will be called the "chart format" in what follows.) Neither the likelihood ratio format nor the chart format is commonly employed at this time in criminal cases in the United States, although they do appear in civil paternity cases and the occasional criminal case in which paternity is material.[166] Most criminal courts, however, allow the presentation of the random match statistic, expressed either as a probability (for example, 0.04) or as a frequency (1 in 25, or 4%).[167] (These conventional presentation methods will be called variations of a "frequencies format.")

---

[166] See 1 David L. Faigman et al., Modern Scientific Evidence: The Law and Science of Expert Testimony §§ 15-5.4 to 15-5.5, at 656–59 (1997).

[167] See 1 id. § 15-5.3, at 652–56.

The question of the best way to present such evidence to a jury is, in significant part, an empirical question. Allen and Leiter concur, discussing research done by Professor Jonathan Koehler on the ability of jurors properly to evaluate evidence about random match probabilities.[168] Interestingly, given Allen and Leiter's orientation, that research does not directly address the question of greatest interest from a veritistic perspective, namely how to present the evidence to a jury so as to obtain the most accurate results.[169] A considerable body of other reported research, not discussed by Allen and Leiter, does address this question. Interestingly enough, it uses a Bayesian benchmark for what an accurate assessment of the evidence would be.[170] This research has generally supported two conclusions relevant here. First, regardless of format used, mock jurors tend to undervalue forensic match evidence,

---

[168] Allen & Leiter, supra note 1, at 1543–45 (discussing Koehler et al., supra note 165).

[169] The results generated by Koehler and his colleagues lend credence to the idea that jurors may need to be controlled by the exclusion of evidence about very small random match probabilities, those that are dominated by much higher rates of false positive lab errors, because jurors seem to be irrationally impressed by small random match probabilities in such contexts. Because of limitations inherent in the research, Allen and Leiter express appropriate caution about the wisdom of an exclusionary rule, id. at 1543–45, but it is easy to understand why Allen and Leiter, who emphasize the jury control principle, would find appealing an interpretation of the results that characterizes the jury as misled by the random match probability. Koehler's results can be interpreted in other ways, however. For example, it may be that including evidence of the random match probability allows the jurors to feel greater confidence in the evidence because it is presented to them in a more complete way. This confidence could translate into assessments of the probability of guilt that are *closer* to what they should be. In other words, rather than jurors being overimpressed by an irrelevant random match probability, it may be that they are underimpressed by a match report when the random match information is suppressed or when no meaningful guidance is provided in the testimony as to how to combine the random match probability with a comparatively large lab error rate. Koehler's analysis does not address this because he does not attempt to determine a normative standard—a standard that would assess where his subjects ought to come out in the case presented.

[170] E.g., David L. Faigman & A. J. Baglioni, Jr., Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence, 12 Law & Hum. Behav. 1 (1988); Jane Goodman, Jurors' Comprehension and Assessment of Probabilistic Evidence, 16 Am. J. Trial Advoc. 361 (1992); Jason Schklar & Shari Seidman Diamond, Juror Reactions to DNA Evidence: Errors and Expectancies, 23 Law & Hum. Behav. 159 (1999); Brian C. Smith et al., Jurors' Use of Probabilistic Evidence, 20 Law & Hum. Behav. 49 (1996); William C. Thompson & Edward L. Schumann, Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy, 11 Law & Hum. Behav. 167 (1987).

relative to the Bayesian norm.[171] Second (a conclusion with considerably less support), no significant reduction in this undervaluation results from instruction to jurors, as in the chart format described above, about how Bayes' Theorem might apply to the case.[172]

Given Allen and Leiter's skepticism about Bayesian analysis, it is not surprising that they do not discuss a body of research that illustrates the value of Bayesian formalism in assessing the use of evidence in courts, including the use of Bayesian explanations in court. Allen and Leiter might find solace, however, in the second of the aforementioned conclusions. It appears to suggest that Bayesian instruction does not speak to jurors in a language that they can understand or accept.[173] Unfortunately for Allen and Leiter, this conclusion is seriously undermined by empirical research that I have been conducting over the last two years. In the rest of this Section, I summarize a portion of the results of that research.[174]

The data for the study were collected in Kane County, Illinois, using the jury pool for that county's criminal court. In one part of that research, 542 jurors called for service were given a hypothetical rape case, the evidence in which was described in writing, and were asked both to assess the probability of the defendant's guilt

---

[171] This proposition is supported by each of the studies cited in the previous note.

[172] See Faigman & Baglioni, supra note 170; Smith et al., supra note 170. With regard to the effect of using what I have called the likelihood ratio format, one study found significant effects on the probability of guilt assessed by subjects but did not attempt a comparison to any normative measure of the posterior probability, whether Bayesian or otherwise. See Jonathan J. Koehler, On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates, 67 U. Colo. L. Rev. 859, 880–83 (1996) (studying, inter alia, a likelihood ratio format that did not also include testimony stating or explaining the random match probability itself).

[173] In a possible reference to this body of research, Allen and Leiter comment in a footnote:

> [I]t is a complete mystery whether DNA evidence can be incorporated algorithmically into trials in a manner that increases the accuracy of decision. "Algorithmically" here is an important qualifier. Obviously DNA evidence can easily be a primary determinant of the truth of competing stories, but for such a purpose no formal analysis of the type discussed here need be employed.

Allen & Leiter, supra note 1, at 1520 n.94.

[174] The full report is given in Dale A. Nance & Scott B. Morris, A Bayesian Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability (Aug. 20, 2001) (unpublished manuscript, on file with the Virginia Law Review Association). It is anticipated that this paper will be published in an upcoming issue of Jurimetrics Journal, together with other proceedings of the Fourth International Conference on Forensic Statistics.

and to indicate what verdict they would give in the case. The subjects were randomly divided into five principal experimental conditions. Some subjects were given no forensics match evidence. Their responses measured the "prior probability" of guilt—the probability that would be assessed without the match evidence. In each of the other four groups, the match evidence was added but presented in one of four different formats. One group was given no statistical information relating to the match. Another group was given testimony reporting both the random match probability—given as 4% and explained as "1 in 25"—and a false positive lab error rate estimate—given as "1 in 1000". In other words, they received the statistical evidence in a frequencies format. Another group received this information as well as testimony translating the four percent random match probability into a 25:1 likelihood ratio—that is, they also received the evidence in the likelihood ratio format.[175] Finally, the last group also received instruction in the effect of a 25:1 likelihood ratio using the chart format provided above.[176] Two distinct measures of the Bayesian normative likelihood ratio, and corresponding posterior probabilities, were calculated from data provided by the subjects regarding their expectations about the chances of a random match, a laboratory error, or some other source of a false positive match report, such as witness perjury or mishandling of specimens by the police.[177]

---

[175] Id. at 13. Notice that for this experiment, the random match probability reported in the testimony is forty times larger than the lab error as so reported. Thus, there was no reason to be concerned that the lab error would rationally dominate the random match probability. See supra note 169. Also observe that in this variant of the likelihood ratio format, the witness does explain the random match probability and how it is related to the likelihood ratio, rather than simply stating the ratio itself. Cf. Koehler, supra note 172, at 880–81 (presenting results of a study in which only the ratio was stated in one condition).

[176] In both the likelihood ratio format and the chart format presentations, the jury was reminded that the likelihood ratio of 25:1 does not take into account the risk of lab error or of other causes of a false finding or reporting of a match. Because the random match probability is so much larger than the testimonial lab error rate, no attempt was made to combine the two figures. See supra note 165.

[177] The methodology for articulating a Bayesian norm has improved significantly over the last fifteen years. In my study, this important step was performed both with and without the assumption that uncontradicted testimony about the random match probability and the lab error rate would be accepted by the jury at face value. The two values were found to be quite comparable, and the conclusions stated in the text do not depend on which measure is chosen.

Using the largest and most jury-representative pool of subjects so far studied in this connection, the results support the first proposition stated above: that jurors tend to undervalue the evidence of the match under all presentation formats examined. The results conflict with the second proposition, however. Whereas earlier studies had found little or no significant improvements from Bayesian instruction (that is, comparing the chart format to the conventional frequencies format), my study found that such instruction closed about sixty percent of the gap between the assessed probability of guilt and the average of the two Bayesian norms, a difference both statistically and practically significant. Use of the likelihood ratio format, without the chart-referencing explanation, yielded intermediate effects, but the differences did not attain statistical significance by conventional standards.[178]

There are, of course, a number of ways that Bayesioskeptics like Allen and Leiter might try to reject these results as unhelpful, but I will not try to anticipate their responses.[179] My tentative interpretation of the experimental results is that the quite rational tendency of jurors to discount the probative value of technical evidence that they do not fully understand is partially offset when the probative value of that evidence is clearly illustrated. If this interpretation is confirmed by further study, then the Bayesian formalism can serve a useful function *in* trials as well as in the assessment *of* trials.

CONCLUSION

As philosophy, the veritistic approach encouraged by a naturalized epistemology is not without its difficulties. To take just one interesting example, in the veritism of Professor Goldman, a belief

---

[178] Further portions of the study assess the extent to which the apparent accuracy gain associated with the chart and likelihood ratio formats can be attributed to fallacious reasoning by the subjects. The results suggest that the so-called "prosecutor's fallacy" identified in previous research, see Thompson & Schumann, supra note 170, which might improperly elevate the assessed probability of guilt, did not occur to any measurable extent under the chart format, but did occur in about 8% of the subjects given evidence under the likelihood ratio format. Some subjects succumbed to other fallacies favoring the defense. This accounts in part for the failure of subjects' average assessments to reach the Bayesian norm under any format.

[179] Various arguments and replies are assayed in the detailed paper. See Nance & Morris, supra note 174.

counts as knowledge if "caused by a generally reliable process."[180] That is laudably empirical in the way it directs our attention, providing a bridge between philosophy and science. But one may well ask, how do we know that any of our cognitive processes are reliable, when to make such an assessment itself requires some reliable cognitive process? Is there not a vicious circularity here?

What is most interesting, in this context, is the manner in which Goldman responds to this concern. In his book, *Knowledge in a Social World*, he devotes several pages to the matter.[181] He invites us to consider the consequence of multiple sense reports, under varying conditions, each reporting the same proposition, such as that there is a peach on the table. What ought we to infer from such observation reports? He notes that the data are consistent with the hypothesis that an evil demon consistently manipulates our perceptions so as to fool us into thinking there is a peach on the table, but he argues that the prior probability of this explanation is low. He notes that the data are (logically) consistent with the hypothesis that sense perceptions might be unsystematically unreliable, but argues that the probability of consistent sense reports then becomes steadily smaller as the number of confirming reports increases. Finally, he notes the consistency of the data with the hypothesis of reliability in the sense perceptions. The astute reader will see where this is going. Goldman then applies Bayes' Theorem, arguing:

> [A] visual corroboration of vision's first report is more likely on the hypothesis that vision is reliable (and both reports are correct) than it is likely on the hypothesis that vision is (unsystematically) unreliable (and both reports are false). Again, this is because an unreliable practice is not so likely to issue in the same mistaken judgment twice (or three or four times). As long as the likelihood of the corroboration is higher given the reliability hypothesis than given the (unsystematic) unreliability hypothesis, corroboration events provide evidence in favor of reliability.[182]

---

[180] Allen & Leiter, supra note 1, at 1494 (quoting Alvin I. Goldman, Epistemology and Cognition 51 (1986)).

[181] See Goldman, supra note 43, § 3.3, at 83–87.

[182] Id. at 86.

Of course, Allen and Leiter can reply that this argument is unrelated to their claims about trials since the inference being made is that of a philosopher thinking about epistemology rather than a jury thinking about a lawsuit. Nonetheless, it is striking how a champion of veritism whom Allen and Leiter repeatedly endorse uses Bayesian reasoning to support the very foundations of reliability in trials—our dependence on the reporting of sense perceptions by witnesses.[183] My point is simply that it is all too easy to underestimate the diversity of ways that such an "algorithmic" approach to probabilities can be useful in understanding the problems of inference at trial.

In the final analysis, I agree with Allen and Leiter that the advent of naturalized epistemology is a happy development in philosophy, at least from the point of view of those who work in the field of evidence law. I agree with Allen and Leiter that philosophical veritism "solidifies the ground beneath their feet." In my view, however, this is true not only of the traditional evidence scholars. It is also true of those who have pushed the theoretical envelope into the domains of expected utility theory and Bayesian probability revision. These, too, are legitimate tools in a veritistic approach to understanding and improving our trial procedures.

---

[183] See, e.g., Fed. R. Evid. 602 (limiting witnesses to testimony about matters about which they have "personal knowledge").