

Penn State Law eLibrary

Journal Articles **Faculty Works**

1986

Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?

David H. Kaye Penn State Law

Follow this and additional works at: http://elibrary.law.psu.edu/fac_works



Part of the Criminal Law Commons, and the Evidence Commons

Recommended Citation

David H. Kaye, Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?, 66 B.U. L. Rev. 657 (1986).

This Article is brought to you for free and open access by the Faculty Works at Penn State Law eLibrary. It has been accepted for inclusion in Journal Articles by an authorized administrator of Penn State Law eLibrary. For more information, please contact ram6023@psu.edu.

COMMENT

DO WE NEED A CALCULUS OF WEIGHT TO UNDERSTAND PROOF BEYOND A REASONABLE DOUBT?†

D. H. KAYE*

Nearly a decade ago, Dr. L. J. Cohen described a system of ordinal, inductive "probabilities" said to measure the strength of certain arguments and said to be very much in the spirit of Sir Francis Bacon. In his current article, The Role of Evidential Weight in Criminal Proof, Cohen explains how his inductive probabilities are closely connected to the work of John Maynard Keynes. Building on some tentative thoughts that Keynes expressed in 1921, Cohen argues that probabilities per se are insufficient to describe the probative force of evidence and that a calculus of "weight" is required to supplement or supplant these probabilites. He asserts that one must attend to the completeness of the evidence used to estimate a probability, and that a measure of completeness cannot be incorporated into the probability itself or by means of another probability. Cohen develops his argument with reference to the burden of persuasion in criminal cases:

[I]n deciding on a course of action, it seems plausible to suppose that we ought to take account of the weight as well as the probability of different expectations. But it is difficult to think of any clear example of this, and I do not feel sure that the theory of "evidential weight" has much practical significance.

^{† © 1986} by D. H. Kaye. All Rights Reserved.

^{*} Professor of Law and Director, Center for the Study of Law, Science and Technology, Arizona State University. Michael White and Laurence Winer provided useful comments on a preliminary draft of this paper. My treatment of this topic was stimulated by conversations with Ronald Allen, Ira Ellman, and Peter Tillers, and refined by observations from the faculty who attended workshops at the University of Pennsylvania and Arizona State University.

¹ L. Cohen, The Probable and the Provable (1977).

² Cohen, The Role of Evidential Weight in Criminal Proof, 66 B.U.L. Rev. 635 (1986).

³ J. KEYNES, A TREATISE ON PROBABILITY 71-78 (1921). Keynes prefaced his discussion of "weight" with the remark that "[t]he question to be raised in this chapter is somewhat novel; after much consideration I remain uncertain as to how much importance to attach to it." *Id.* at 71. Keynes further observed:

Id. at 78. Cohen, it appears, does not share these doubts.

[I]f you try to give an account of the standard of criminal proof in Pascalian terms—i.e., as requiring a very high value for a probability function that conforms to the axioms of the mathematical calculus of chance—you are inevitably driven to reserve the crucial place in your reasoning for the assignment of a high value to a non-Pascalian function for the assessment of evidential weight.⁴

I believe that Cohen's fundamental premise is correct. One must examine the completeness of a body of evidence and the circumstances under which the evidence was gathered if one is to assess its probative value. But I disagree with Cohen's conclusion that one is "inevitably driven" to construct a model of forensic proof that relies on a conception of the "weight" of a body of evidence that is logically distinct from the probability of an event conditioned on that body of evidence. Nor is it apparent to me that Cohen's method of expressing "weight" represents an advance over an analysis that relies on conventional, cardinal probabilities.

To establish that resort to a non-probabilistic measure of weight is unnecessary, we can criticize the details of Cohen's reasoning and comment on his examples. Alternatively, we can simply exhibit a model of forensic proof that (a) relies exclusively on conventional probability theory, and (b) accomplishes the tasks that Cohen thinks must be borne by an independent concept of evidential weight. I shall do a little of each, but mostly I shall try to pursue the second route. That is, I shall sketch an account of courtroom inference based strictly on probabilities and an interpretation of the burden of persuasion derived from decision theory.

If my elaboration on this conventional interpretation of the burden of persuasion is not flawed in the ways that Cohen believes any strictly probabilistic analysis must be, then his latest complaint against standard probability theory must fail.⁵ Part I of this paper sketches the basic outlines of the conventional probabilistic interpretation of the burden of persuasion. Part II supplements this theory to allow it to recognize gaps in the state's case. Part III offers a few observations about a separate problem that Cohen mentions,

⁴ Cohen, *supra* note 2.

⁵ More precisely, his claim that he demonstrated that reference to non-probabilistic weight is necessary to an explanation of the burden of persuasion in criminal cases will have been shown to be false. This is not to say that a theory that does use non-probabilistic weight is necessarily implausible or inferior to a totally probabilistic theory. I do not deal with the question of comparative merit in any depth, however, because I do not feel qualified to judge whether Cohen correctly elucidates Keynes's idea of evidentiary weight or whether Cohen's inductive probabilities solve the problem he perceives. Cf. Allen, A Reconceptualization of Civil Trials, 66 B.U.L. Rev. 401 (1986) (questioning the capacity of Cohen's calculus to resolve other "paradoxes").

namely, the failure of the decision-theoretic analysis to articulate interpersonal standards for arriving at subjective probablities.

I. THE CONVENTIONAL INTERPRETATION OF THE BURDEN OF PERSUASION

A. Evidence, Contentions, and Stories

To establish that the conventional decision-theoretic interpretation of the burden of persuasion is a viable theory, I must be more comprehensive and precise than most writers—myself included—have been. The detail is tedious, yet it is necessary to avoid confusion. For clarity, I shall use some symbols, but I have no formal proofs to give. Thus, while the analysis is not mathematically rigorous, it has the advantage of being accessible to any moderately determined reader.⁶

At the outset, I want to distinguish between "evidence," "factual contentions," and "stories." An item of evidence is a stipulation, a fact established by judicial notice, or most frequently, the portion of the testimony of a witness or a documentary exhibit that is introduced to help prove a factual contention. With the exception of matters of which judicial notice is taken, evidence comes from two sources: the prosecution and the defense. The

Although the definition of probability as a function whose domain is a sample space can be found in virtually all mathematics texts, logicians often speak of probability as being distributed over sentences in a formal language. E.g., R. Jeffrey, The Logic of Decision (2d ed. 1983). If my analysis were to be recast in these terms, the same conclusions would obtain. Cf. P. Suppes, Probabilistic Metaphysics 187 (1984) ("[T]here is a classical tradition of ambiguity in probability theory of whether to speak of the probability of events or of propositions, and I do not think it important which one we choose.").

⁷ Professor Nesson's brief but provocative treatment of the conjunction of evidence as opposed to the conjunction of elements of a cause of action prompted me to think in these terms. See Nesson, The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts, 98 HARV. L. REV. 1357, 1387-90 (1985). Cohen does not attend to such distinctions. He speaks of "facts" sometimes to mean the evidence before the court, and sometimes to mean factual contentions that are accepted as true.

⁶ I use some symbols that are standard in set theory. These include the curly brackets "{" and "}" to indicate that the items listed between the brackets form a set, "∪" to denote the union of two sets, and "∩" to indicate the intersection of two sets. The elements of some of the sets that I describe are events in a sample space. The symbols referring to sets that are within the sample space are italicized. The probabilities to which I refer are values of a non-negative, additive function, whose maximum value is one, defined on this sample space. I belabor these points because of Cohen's concern that the probability function that I use to measure rational, partial beliefs may lack "appropriate semantics." Cohen, supra note 2, at 646 n.19.

prosecution offers some number n of "items" or "pieces" of evidence (assertions of a witness, photographs, etc.), and the defense presents m items of its own. The prosecution's evidence can be represented as $\{E_1^P, E_2^P, \ldots, E_n^P\}$. The defense's evidence can be represented as $\{E_1^D, E_2^D, \ldots, E_m^D\}$, and the totality of the evidence is $E^P \cup E^D$.8 The jth item of prosecution evidence may be denoted E_j^P , where j is between one and n. E_j^D has similar meaning.

If we think of the trial as analogous in some respects to an experiment⁹ with various possible outcomes (events) whose probabilities can be computed (under various hypotheses) prior to running the experiment, then the presentation of the pieces of evidence comprising E are like the events (the experimental data) that the experimenter observes. E_1^P is the event that the prosecution presents E_1^P (the jth item of evidence), E^P is the event that the prosecution offers the items of evidence E_1^P through E_n^P , and so on. In other words, E^P , E^D , and the joint event E (which is the intersection of these two events) states what happened at trial, including features such as the identity of the party adducing the evidence and the demeanor of a witness, that may bear on the credibility of the propositions being asserted.

In an armed robbery case, for instance, E_1^P might be testimony of a store clerk that the defendant is the woman who robbed the store, E_2^P might be his testimony that she pointed a gun at him, E_3^P might be a police officer's testimony that he found a gun in the defendant's apartment, and so on. Of course, the defendant may also produce evidence. E_1^D might be the testimony of the defendant's mother that she and her accused daughter were at a movie theater watching *Gone with the Wind* during the time of the robbery, and the final item of evidence might be E_9^D , the testimony of defendant's minister to the effect that the defendant has a good reputation for being law-abiding and peaceable.

The evidentiary events differ from the outcomes of an experiment in many ways. For one, an impassive Mother Nature who determines the experimental results does not dictate the evidentiary outcomes. Rather, the parties introduce the items of evidence to convince the judge or jury of their factual contentions. These contentions are statements about what happened in the past, and we may think of them as events to which probabilities can be assigned, just as we described the presentation of evidentiary items as events transpiring at trial. The prosecution may contend, for example, that the defendant fled from the store in her car. Each such factual contention advanced by the prosecution can be denoted C_j^p , where, as with the items of evidence, j indexes each such contention. We may let r represent the total number of these contentions of the prosecution. Analogously, the defense's

⁸ Where the individual pieces of evidence are distinct, this union is $E = \{E_1^P, \ldots, E_n^P, \ldots, E_1^D, \ldots, E_m^D\}$.

 $^{^9}$ See I. Good, Good Thinking: The Foundations of Probability and Its Applications 11-12 (1983).

contentions of fact may be denoted C_j^D , and there are some number s of these.

The number of contentions, r + s, and the number of items of evidence, n + m, need not be equal. A single item of evidence may bear on several contentions. Alternatively, several items of evidence may bear on a single contention. In addition, some of the defense contentions may be simple negations of the plaintiff's contentions, while other contentions may be identical.

The totality of the factual contentions presented by each side make up the "stories" that the parties tell. S^{D} is defendant's story. It consists of the joint event C_{1}^{D} through C_{n}^{D} . If the defendant declines to make an opening statement and moves for a directed verdict at the close of the prosecution's case, S^{D} is not yet known.

 S^P is the prosecution's story. It is the joint event C_1^P through C_r^P . For there to be a disputed issue of fact, S^P and S^D must differ as to at least one of their constituent contentions. Furthermore, because the stories are constructed in light of the pertinent law, S^P is such that all the elements of the offense are satisfied, while S^D is such that one or more elements are not satisfied.¹⁰

B. The Posterior Probability of a Story and the Burden of Persuasion

The conventional account of the burden of persuasion supposes that factfinders review all of the evidence and, aided by closing arguments and relying upon their conception of human behavior and the way the world works, somehow arrive at a personal probability that plaintiff's story, S^{P} , is true, as opposed to the defendant's story, S^{D} . This personal probability can be represented as

$$\Pr(S^{P}|E) = p \tag{1}$$

The vertical bar between S^P and E emphasizes that the probability of S^P is conditioned on the receipt of all of the evidence E. The factfinder is not allowed to rely on extra-judicial knowledge of the case. Because the proba-

¹⁰ 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.

The stories play the role of hypotheses in our analogy to a statistical experiment. Since we will want to speak of conditional probabilities for the competing stories, the mathematical formalism requires that they be treated as constituting a set analogous to the parameter space for such an experiment.

¹¹ See, e.g., Mansfield, Jury Notice, 74 GEO. L.J. 395 (1985).

bility p is formed after all the evidence is in, we may call it a posterior probability.¹²

The burden of persuasion is then the simple rule that the judge or jury should convict or render a judgment for plaintiff if and only if p > p,* where p* is some critical probability (some number close to one in criminal cases). Deciding cases according to this criterion minimizes expected losses, where the loss function is a linear function of the numbers of false convictions and false acquittals.¹³ This model can be used to describe and evaluate evidentiary rules concerning the burden of persuasion and presumptions or inferences.¹⁴

II. THE POSTERIOR PROBABILITY AND "GAPS" IN THE EVIDENCE

Cohen's chief criticism of this account is that p is a posterior probability conditioned on a body of evidence E that may be incomplete or misleading. According to Cohen, the dispositive probability must be unconditional. In other words, it is necessary to move from the posterior probability $Pr(S^P|E)$ to an unconditional probability $Pr(S^P)$. While this phrasing is misguided, ¹⁵

¹² Also, because we are considering the normative question of how jurors ought to assess evidence, we do not deal with the personal probabilities that might be entertained by real jurors affected by passion, prejudice or misinformation, but with the probabilities that ideal jurors would come to have on the basis of the evidence E. See Lempert, Modeling Relevance, 75 Mich. L. Rev. 1021 (1977). This prescriptive model can serve as a standard for evaluating the likely behavior of real jurors and for constructing rules of evidence that (to the extent desired) will promote accurate assessment of the probabilities. The problem of interpreting the personal probabilities is considered further in Part III.

¹³ For a discussion of the difficulty of defining this loss function precisely, see Kaye, And Then There Were Twelve: Statistical Reasoning, the Supreme Court, and the Size of the Jury, 68 CALIF. L. Rev. 1004 (1980). It is also important to recognize that the decision-theoretic interpretation does not mean that actual losses necessarily are minimized. Rather, the rule is constructed to minimize the expectation of the loss function. The law of large numbers establishes the link between expected values and long run frequencies. Not all commentators are careful to distinguish between expected values and actual outcomes. E.g., Nesson, supra note 7, at 1377 n.67.

¹⁴ E.g., Kaye, The Limits of the Preponderence of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation, 1982 Am. B. FOUND. Res. J. 487; Kaye, Probability Theory Meets Res Ipsa Loquitur, 77 MICH. L. Rev. 1456 (1979); Kornstein, A Bayesian Model of Harmless Error, 5 J. LEGAL STUD. 121 (1976); Lempert, supra note 12; Note, A Probabilistic Analysis of the Doctrine of Mutuality of Collateral Estoppel, 76 MICH. L. Rev. 612 (1978).

¹⁵ Cohen urges that "we speak of deriving an unconditional probability . . . because that seems closer to the actual language and purpose of the criminal courts." Cohen, *supra* note 2, at 638. It is not clear to me what he has in mind. The "actual language" of the courts usually includes an instruction to the jury to decide the case on the strength of the evidence admitted. This does not suggest an unconditional

the recognition that $Pr(S^p|E)$ does not contain information about the completeness of E is an important insight. Either E must be interpreted differently or a different equation must be substituted for (1). This can be done within a framework of conditional probability without resorting to a distinct calculus of evidentiary weight.

A. The Prevalence of Gaps

The exposition of the burden of persuasion in Part I ignores the possibility of gaps in a litigant's evidence—gaps that make the party's story less believable. Any good trial lawyer knows that the jury will expect to hear certain items of evidence in certain cases, and that it may regard the failure to produce such evidence with devastating skepticism. ¹⁶ Let me offer an

probability. As for the "purpose of the criminal courts," that is not one, but many things. It is not obvious that basing decisions on unconditional probabilities would better promote these functions than returning verdicts on the basis of conditional probabilities. This phrasing is not meant to suggest that I advocate instructing jurors to articulate any kind of a probability as an aid to decisionmaking.

Perhaps my confusion arises from an idiosyncratic use of the phrase "unconditional probability." There are indications in Cohen's paper that he does not use the unconditional probability of an event A to represent the probability of A prior to the receipt of a body of evidence E relevant to A, but rather that he means to indicate the probability of A conditioned on all conceivable relevant evidence about A (the "totality of the facts"). If so, and if we let T stand for the presentation of this totality, then the "unconditional" probability of which he speaks is P(A|T). Of course, we could omit the explicit reference to T, and merely write P(A), but this abbreviated notation would not make the probability of A, assessed in light of T, into an unconditional probability. Cf. J. Keynes, supra note 3, at 40 (remarking on the errors that arise when notation fails to make the conditional character of a probability explicit); id. at 119 (urging the adoption of notation that makes the premises of an argument explicit to avoid "endless confusion" about probability).

¹⁶ My former colleague, Wendell Kay, an eminent trial lawyer, used to tell his students of a news reporter's interview with a juror in 1979, reported in an Associated Press story, that indicates that the ordinary scenario, as the jury commences its deliberations, goes something like this:

Juror #1: That guy looked guilty to me. Did you ever see such shifty eyes, and did you notice the way he slouched at the table?

Juror #2: Well, I don't think that has anything to do with it, but you know darn well that the police must have had something to go on or they never would have arrested the guy in the first place.

Juror #3: Well, the Judge told us not to pay any attention to the fact that he got indicted, but I sure say that I never saw a cloud of smoke without some fire burning somewhere.

Juror #4: When do we go to lunch?

Juror #5: The guy sounded pretty good to me, but if this fellow Wilson was actually with him like he said, why didn't he call Wilson as a witness?

anecdote drawn from my wife's experience as a juror. The state charged a young man with driving while intoxicated (DWI). A police officer testified that he stopped the defendant after seeing him change lanes illegally at two a.m. He further testified that he smelled alcohol on the man's breath, and that a "field sobriety test" showed the defendant to be intoxicated. The defendant's story was that he was lost in an unfamiliar neighborhood, listening to his wife in the car and trying to determine which way to proceed. The jury believed that the defendant probably was drunk, but it acquitted because the state, which had to prove guilt beyond a reasonable doubt, did not measure the defendant's blood or breath alcohol level. The jury expected to hear "breathalyzer" evidence in a drunk driving case, and the absence of that evidence in the state's case raised a reasonable doubt in the minds of the jurors.¹⁷

These "negative inferences" extend to civil matters as well. Consider a paternity case in which the plaintiff concedes that two men could have been the father. Suppose the plaintiff compels the defendant to submit to immunogenetic testing, and inexplicably ignores the other man. Even if the genetic tests implicate the defendant, the plaintiff's story is weaker than it would be if both men had been tested and the nonaccused man excluded as a potential father. 18

Kay, Final Argument (unpublished manuscript). The remark of Juror 5 is the type that this comment tries to analyze.

¹⁷ There are two distinct reasons why the jury might have reached this result. One is that it actually found the other evidence less persuasive. The "negative inference" would be that the police officer decided not to collect scientific evidence because he wanted to issue a citation but his "field sobriety test" was really not so conclusive, and hence he feared that the more scientific tests might have exonerated the defendant. The other possibility is that the jury decided to "punish" the police for being "lax," that is, for not following procedures that the jury felt the state should follow in dealing with citizens suspected of drunk driving.

In previous articles, I have emphasized the latter systemic concern as a justification for some putative rules of law that preclude liability on the strength of evidence that seems to satisfy the pertinent burden of persuasion. Kaye, The Laws of Probability and the Law of the Land, 47 U. Chi. L. Rev. 34 (1979); Kaye, Paradoxes, Gedanken Experiments and the Burden of Proof: A Response to Dr. Cohen's Reply, 1981 Ariz. St. L.J. 635. Despite my window into the jury room, I cannot say whether the jury in this case declined to convict because it drew a "negative inference," or because it adopted a systemic perspective of the sort that might explain certain common law rules, or both. The jurors did not articulate why they found the gap in the state's case so troubling.

¹⁸ Cf. Ellman & Kaye, Probability and Proof: Can HLA and Blood Testing Prove Paternity? 54 N.Y.U. L. Rev. 1131, 1158-61 (1979) (advocating a rule of law that would bar plaintiffs from recovering when they choose not to introduce more revealing evidence which is reasonably available to them). For examples of judicial recognition of the impact of the negative inference resulting from a gap in the case, see Galloway v. United States, 319 U.S. 372, 385-88 (1942); Hirst v. Gertzen, 676 F.2d 1252, 1259-60 (9th Cir. 1982).

Any model of courtroom reasoning that cannot handle the negative inferences that arise from the obvious incompleteness of a party's evidence is deficient. As I have indicated, the crucial question is whether our probabilistic analysis of the burden of persuasion can be modified to reflect valid reasoning about gaps in a party's case.

B. A Probabilistic Account of Evidence in the Context of a Story

Evidence cannot be evaluated in a contextual vacuum. It is puzzling to read of "John Doe" holding a losing lottery ticket without being told how the "lottery" was conducted. ¹⁹ In the forensic arena, a party's evidence is

¹⁹ I am referring to the following hypothetical case presented by Professor Cohen: Imagine a lottery with as many tickets as you please and only one winning ticket. Then the proposition that John Doe's ticket is not the winning one has as high a conditional probability as you please on the evidence available about the total number of tickets participating. But you would certainly be rash to infer from this that to a very high *un*conditional probability John Doe's ticket is a loser. You would still have some reason to doubt the validity of the inference until you know that the lottery was run fairly. For example, if it were also a fact that John Doe is a mobster whose gunmen supervise the draw, that fact would be part of the totality of facts that a reasonable person would think relevant to your conclusion. Consequently the probability of your conclusion on the basis of the totality would be a good deal lower than its probability on the evidence that was originally available.

Cohen, supra note 2, at 638.

For clarity, let us define some terms. At the outset, there are three facts to consider: that John Doe holds a lottery ticket, that there are N tickets, and that exactly one is the winning ticket. Let L be the event that Doe's ticket is a loser. The probability of L then can be denoted Pr(L). To estimate this probability, we need to adopt some model of how Doe acquired his ticket. The three background facts are not part of any sample space needed to define the probability that Doe holds a losing ticket. They merely bear on the probability model that gives the unconditional probability Pr(L).

Evidently, Dr. Cohen believes that in the absence of any more information about the distribution of tickets, one should presume that each ticket is equally likely to be the winner. On this assumption of random sampling from a finite population, the probability distribution is uniform, and Pr(L) = (N-1)/N. But the validity of the random sampling model is an empirical, not a logical matter, and we must consider other plausible models before embracing (N-1)/N as the probability in question. Perhaps John Doe is a lottery official who knows the winning ticket in advance and has appropriated it for himself. Perhaps he is the mobster who secretly controls the lottery, as Dr. Cohen posits. In the context of such models, the probability Pr(L) would seem to be zero, or close to it.

If we must estimate Pr(L) without learning more about the model that actually pertains, then we must rely on our background knowledge about the prevalence of fair lotteries and about John Doe's character. If all lotteries are fair, we should take Pr(L) to be (N-1)/N. If all are rigged for John Doe, then Pr(L) approaches 0. Our best guess probably would lie somewhere between these extremes. In sum, a

evaluated in the context of the whole story that the party is presenting. The prosecution's story S^P will suggest that various items of evidence will exist if the state has been diligent in preparing its case. The jury in the DWI case, for instance, believed that if the state's story had been true and if the police had been diligent, then evidence of the concentration of alcohol in the defendant's breath would have been presented.

More generally, we may denote the collection of evidence that one might expect to hear in support of S^P (if S^P is what occurred) as $F^P = \{F_1^P, F_2^P, \ldots, F_t^P\}$. Also, let us use F^D to stand for the collection of evidence that one would expect to hear in support of S^D if S^D occurred. In our probabilistic model of the fact-finding process, we may say that the judge or jury should contrast E^P with F^P , and E^D with F^D . If F_j^P are missing from E^P , then these gaps G^P in the state's case are part and parcel of the event E conditioning S^P . A gap is as real as a "hole" in a semiconductor. We might say that E refers not to the presentation of evidence in a vacuum, but to evidence in context, that is, with G^P (and G^D —the gaps, if any in defendant's evidence). It will be convenient to let G stand for all these gaps $G = G^P \cup G^D$ and to let G be the corresponding joint event that there were these evidentiary gaps G at the trial.

Then we can make the full meaning of the joint event E explicit by substituting for $\Pr(S^P|E)$ in (1) the expression $\Pr[S^P|E\cap G)$. In other words, the decision rule becomes: Find for the defendant if and only if

$$\Pr[S^{p}|(E \cap G)] = p < p^{*}$$
(2)

Given the importance of perceived gaps in a party's evidence, the rule of evidence should afford litigants the opportunity to introduce evidence and to explain to the jury why there appear to be gaps in their cases, assuming that this does not implicate other important values.²⁰ If G^P is not explained

thoughtful statistician or probabilist would not maintain that Pr(L) must be (N-1)/N, for he or she would realize that this value presupposes the random sampling model, which may not hold for the hypothetical lottery. The facts that Dr. Cohen initially enumerates must be viewed in some context.

This is not to deny the main point of Dr. Cohen's hypothetical crooked lottery case. An estimate of Pr(L) derived from a probability model that has very few facts to give it structure may be unstable in the sense that learning more facts could cause us to change our model, and hence, our estimate.

²⁰ See Saltzburg, A Special Apsect of Relevance: Countering Negative Inferences Associated with the Absence of Evidence, 66 Calif. L. Rev. 1011 (1978). Sometimes evidence will be unavailable due to the passage of time or for other innocent reasons. Sometimes evidence will be unavailable because it is privileged or the witness who has it is immune from appearing or testifying. Neither the physically unavailable nor the legally unavailable evidence should diminish the probability of a story.

Inasmuch as Keynesian weight relates to the "amount" of evidence, J. KEYNES, supra note 3, at 74, it might seem as if a "weight-based" theory leads to a contrary and unacceptable result. Yet, Keynes clearly restricts the computation of weight to

satisfactorily, then at the close of the prosecution's case, $\Pr[S^P | (E \cap G)]$ will be less than $\Pr(S^P | E)$. If G^P is a significant enough omission, then this difference may bring p below the critical value p*, precluding a conviction. Instead of turning to a separate calculation of "weight," we account for the completeness of the prosecution's case in the posterior probability itself.²²

"relevant evidence." *Id.* at 71-72. Since he defines relevant evidence as that which alters the prior probability, *id.* at 55, it is tempting to argue that physically or legally unavailable evidence is irrelevant and does not count in the computation of weight.

Still, it is not clear that this defense of Keynes's theory works. His definition of relevance seems to distinguish between a collection of evidence that is irrelevant as a whole but is nevertheless individually relevant. The former adds to weight, but the latter does not. As Glenn Shafer suggested in his remarks at the symposium, perhaps the weight can be computed as the sum of the absolute values of the log-likelihood ratios for each evidentiary event, although this measure may have some technical difficulties. Here the existence of gap G reduces the prior probability (has a negative log-likelihood ratio), but the subsequent evidence of unavailability (having a positive log-likelihood ratio of equal magnitude) boosts us back to where we started. This should add to the weight.

If so, reliance on "weight" does not seem to explicate correct reasoning. Take the case of an alleged purse snatching in New York City in which S^p includes the event that the victim's husband was beside her when a juvenile grabbed the purse and ran straight into the arms of a nearby policeman. The state does not call the husband to testify, and the jury may wonder why this gap exists. So the state establishes that this witness is unavailable because he is the United Nations ambassador from Iran, who has claimed diplomatic immunity. Now consider a similar case in which the woman was walking without her husband. I would think that $Pr(S^p|E \cap G)$ is the same in both cases and that a juror's partial belief in S^p also should be identical. Even though the Keynesian weight of S^p given the evidentiary events in the first case looks like it might be greater than the second, the requirement of proof beyond a reasonable doubt seems satisfied (or not satisfied) to the same degree.

Either I misunderstand Keynesian weight—which is quite possible—or, like Keynes himself, I have great difficulty imagining an instance in which this weight has much practical significance. In this regard, it may be worth noting that Cohen also gives no clear example of a case in which attending to Keynesian weight would warrant a different action than reliance on a well-founded probability.

²¹ Cf. Lindley & Eggleston, The Problem of Missing Evidence, 99 LAW Q. Rev. 86 (1983) (Bayesian analysis).

Michael White has suggested to me that one can account for the completeness of the evidence with what I would construe as a second-order probability. The idea is to obtain an appropriate measure of the invariance of $Pr(S^p|E)$ under conditionalization on F. We want to know whether $Pr(S^p|a)$ is some number a. If the probability would not change much when S^p is conditioned not merely on E but on E together with the presentation of every subset of the additional items of evidence in F, then we should take a number very close to one to express the probability of $Pr(S^p|E)$ being a. Cf. B. Skyrms, Causal Necessity: A Pragmatic Investigation of the Necessity of Laws 12-13 (1980) (defining the "resiliency" of a conditional probability). This line

III. ESTIMATING $Pr[S^P|(E \cap G)]$

The decision rule summarized in (2) responds to Cohen's principal objection to a probabilistic interpretation of proof beyond a reasonable doubt. In the penultimate section of his paper, however, Cohen questions on other grounds the use of personal probabilities in a normative model of forensic proof. Previously, he insisted that such probabilities were wholly inadmissible, since they presupposed a hypothetical lottery or wagering procedure that was meaningless inasmuch as the bets could not be settled and the personal probabilities elicited in this fashion would be affected by the stakes of the bets.²³ In his current article, he at last abandons (or retreats from) these arguments. The view that Cohen once dismissed as "grossly fallacious" is now "easy enough to defend," although he continues to complain that the very idea of determining the personal probabilities of real jurors is "too sophisticated" and "too frivolous." ²⁶

In this regard, whether eliciting personal probabilities with the aid of a hypothetical wagering scenario is a plausible description of a juror's "own state of mind" or of "imaginable introspections of jurors and advocates" is beside the point. No one imagines that jurors see themselves as stating their indifference points for wagers of indeterminate stakes. The wagering formu-

of inquiry leads us into the literature on second-order probabilities. Although there is much work to be done in this field, it now appears that the arguments of the advocates of subjective probability about the coherence of ordinary beliefs can be generalized to demonstrate the coherence of higher-order beliefs. Skyrms, *Higher Order Degrees of Belief*, in Prospects for Pragmatism 109 (D. Mellor ed. 1980).

I mention this approach to the problem because I think it could lead to an interpretation of the burden of persuasion that supplements $Pr(S^p|E)$ (as first presented in (1)) with a fully probabilistic measure of the weight of E. Second-order probabilites also might be capable of capturing the notion of subjective confidence that Professor Neil Cohen wrongly ascribes to a frequentist confidence interval. Cohen, Confidence in Probability: Burden of Persuasion in a World of Imperfect Knowledge, 60 N.Y.U. L. Rev. 385 (1985). For a more careful explanation of the meaning of a confidence interval, see Kaye, Mixing Apples and Oranges: Confidence Coefficients and Significance Levels Versus Posterior Probability and the Burden of Persuasion (1986) (unpublished manuscript).

I shall not pursue the analysis of second-order probabilities here, in part because I lack the technical expertise, and in part because the first-order model that I have sketched should give the same results. See Skyrms, supra, at 167-68 ("first-order models are good medicine for acrophobia").

²³ L. COHEN, supra note 1, at 90; Cohen, Subjective Probability and the Paradox of the Gatecrasher, 1981 ARIZ. St. L.J. 627, 629-32. For a previous rejoinder on this topic, see Kaye, Paradoxes, Gedanken Experiments and the Burden of Proof: A Response to Dr. Cohen's Reply, 1981 ARIZ. St. L.J. 635.

²⁴ L. Cohen, supra note 1, at 90.

²⁵ Cohen, supra note 2, at 646.

²⁶ Id. at 647.

²⁷ Id.

lation of subjective probability is merely one way to motivate the axioms for subjective probability.²⁸ A suitably contrived betting procedure, like the ubiquitous meter rods and clocks of special relativity, is but an idealized measuring device, constructed to induce a certain type of consistency in hypothetical psychological judgments.²⁹

The inquiry, as Cohen finally recognizes, is prescriptive rather than descriptive. He seeks a model that specifies how jurors *should* reason and act. The decision-theoretic analysis focuses on what actions should be taken under risk. It presumes that action must be taken (a verdict must be returned), and it prescribes a rule for doing so that minimizes the expected losses from erroneous verdicts. To this extent, it has normative content. For instance, a juror who sincerely determines that, as far as he or she can tell, $Pr[S^P|(E \cap G)]$ is only one-half, would be unwarranted in voting to convict.

At the same time, the decision-theoretic analysis offered here says much less about belief formation. As Cohen now seems to admit, probabilities can serve to quantify partial beliefs, but they are only the raw material in an idealized scheme of decisionmaking. Consequently, the treatment of the burden of persuasion encapsulated in (2) hardly provides a full account of how jurors should evaluate evidence. For one thing, I have not discussed how p itself should be estimated. I have merely said that if judges and juries are to minimize an appropriate expected loss function, then they must evaluate the evidence to find the probability p and contrast it with p* to reach a verdict. And I have argued that the p in (2) can and should reflect the completeness of each side's evidence.

I have not offered a "Bayesian" interpretation of the burden of persuasion, if by "Bayesian" one means probabilities deduced through Bayes's rule.³⁰ The p in (2) might come from an entirely different reasoning process.

²⁸ See Skyrms, Higher Order Degrees of Belief, supra note 22, at 115-16.

²⁹ Quantifying partial beliefs by reference to what a person would say if he had to decide under a veil of ignorance about which side of a bet of unknown stakes he will be forced to take is an idealization. The procedure is even more hypothetical if the bet cannot be settled, like a bet on the existence of an afterlife. But even this extreme case (more extreme, I think, than the notion of betting on the accuracy of a verdict in a criminal case) does not preclude asking an individual to contemplate the fair odds for such a bet on the strained assumption that a definitive answer to the question of an afterlife will emerge soon. Cf. I. HACKING, LOGIC OF STATISTICAL INFERENCE 215-16 (1965) (acknowledging but questioning the possibility that "you might even imagine yourself pretending to bet on theories and hypotheses, as if you were gambling with an omniscient being who always pays up if you are right").

³⁰ The criterion of minimizing expected loss is called a "Bayes risk criterion." J. Melsa & D. Cohn, Decision and Estimation Theory 44 (1978). In the context of statistical hypothesis testing, a procedure that meets this criterion is called a "Bayes test procedure." M. Degroot, Probabability and Statistics 382 (1975). The phrase "Bayesian" is used in so many ways in the statistical and philosophical literature that it is hard for the uninitiated to know what it means. Kaye, *The Laws of*

For instance, it might reflect an intuitive judgment of the totality of the evidence E assessed in the context of SP and SD, as made with the aid of the various heuristics identified by cognitive psychologists.³¹ Wherever it comes from, as long as the jury's value for p correctly expresses the probability that

S^P is true given the evidence and its context, a decisionmaker who follows the rule given by (2) does the best that we can expect in the risky world of litigation.

But it is clear that Cohen, even if he were to agree that this account obviates the need for an independent measure of evidential weight, would not be satisfied. How do we know whether the judge or jury has arrived at the correct value for p? The analysis thus far gives no insight into the way in which judges or jurors should estimate this crucial quantity. It expresses this probability as a conditional probability—one determined in light of the entirety of the evidence E and its context—but it says nothing about how this conditional proability should be formed. A Bayesian explication of how a factfinder should successively revise an initial prior probability to arrive at p would give the model more structure, but it would merely push the objection one level down.32

Yet, Cohen slides from the legitimate perception that the existing theory is not all-inclusive to the much more problematic contention that "a subjectivist interpretation is inherently incapable" of elucidating "what reasonings are appropriate to jurors and advocates." According to Cohen, "[s]ubjectivism provides us with a way of putting guesses into mathematical uniform, but denies us a way to tell good guesses from bad ones."34

These are puzzling if not disappointing claims. The transition from the statement that "theory T does not provide us with X" to "theory T denies us X" is obviously unsound. Knowing how to solve partial differential equations does not tell a physicist how to arrive at the equations that model the motion of fluid in a container, but the mathematical theory of partial differential equations certainly does not deny him the physical theory needed to discern the correct set of equations and their boundary conditions. Similarly, knowing that idealized jurors can express partial beliefs as conditional prob-

Probability and the Law of the Land, supra note 17, at 45 n.37, 51 n.57. Thus, one could say that my treatment of the burden of persuasion adopts "Bayesian decision theory" but not necessarily "Bayesian learning theory." E. EELLS, RATIONAL DECISION AND CAUSALITY 12 (1982).

³¹ E.g., Judgment Under Uncertainty: Heuristics and Biases (D. Kahneman, P. Slovic & A. Tversky ed. 1982); R. Nisbett & L. Ross, Human Inference: Strate-GIES AND SHORTCOMINGS OF SOCIAL JUDGMENT 117, 241-45 (1980). For that matter, in the abstract at least, it could even be a relative frequency figure, although there is no obvious reference class for the events in question at a trial.

 $^{^{32}}$ See I. Good, Probability and the Weighting of Evidence v (1950).

³³ Cohen, supra note 2, at 647.

³⁴ Id. at 648.

abilities does not bar us from discussing or understanding the epistemology that governs these probabilities.³⁵

Although it is plainly wrong to think that a theory is unsuitable merely because it is not all-inclusive, the question whether the jury's partial beliefs can, in principle, be quantified so that they lead to a suitable probability p in (2) is by no means trivial. Many probability theorists within the personalist or subjectivist school treat personal probabilities as Cohen does. They hold that the degrees of belief quantifed as probability judgments need only obey certain axioms concerning preferences.³⁶ This formal property of coherence may be a necessary feature of a satifactory normative theory of forensic proof, but I agree with Cohen (and many other philosophers) that we ought to demand more than this. Attorneys, courts and jurors do not and should not act as if one sincere probability is as good as another. They worry and argue about what warrants a given degree of belief, which is why a major part of the typical closing statements of counsel consist of exhortations to common sense and generalizations about human behavior and credibility.

It bears reiterating, however, that the use of subjective probability in an idealized theory of forensic proof does not preclude the attempt to formulate a philosophically adequate account of the interpersonal and logical standards that promote accurate estimation of the probability on which, I have argued, a verdict should turn.³⁷ Indeed, Cohen's doctrinaire claim that subjective probability theory is inherently incapable of elucidating such standards dismisses all too cavalierly the efforts of the many probability theorists and students of inductive logic who have attempted to systematize personal judgments of probability in ways that could elucidate such standards.³⁸

³⁵ It also does not suggest that the best epistemological theory includes Cohen's original measure of the strength of an inductive argument.

³⁶ E.g., de Finetti, Foresight: Its Logical Laws, Its Subjective Sources, in Studies in Subjective Probability 97, 111-18 (H. Kyburg & H. Smokler ed. 1964); Ramsey, Truth and Probability, in Studies in Subjective Probability, 63, 69-82 (H. Kyburg & H. Smokler ed. 1964); L. Savage, Foundations of Statistics 27-40 (1954).

³⁷ Cf. D. Mellor, The Matter of Chance xii (1971) (remarking with respect to his propensity theory of physical probability that "nothing in the characterization and measurement of partial belief excludes further empirical constraints of rationality upon it"). According to Mellor, "[p]ersonalists have admittedly tended to accompany their theories with a view of probability statements as merely subjective. But that is an incidental defect of personalism." *Id.* at 2. *See also* Skyrms, *supra* note 22, at 20 (using the phrase "epistemic probability" rather than subjective probability to denote partial belief, and observing that "constraints of rationality no doubt require a good deal more [than coherence]; to say what more is the chief business of epistemology").

³⁸ F. Benenson, Probability, Objectivity and Evidence 44-45 (1984) (distinguishing between "wholly subjective theory," "logical relation theory," and "credal probability" theory). I do not feel qualified to judge which of these approaches is the most promising, and I continue to be agnostic about which camp of philosophers or logicians has discovered the "real" nature of probability.

While it is possible that all attempts to explicate the logical or empirical bases for justifiable subjective probabilities will fail, I think it is too early to foresake this line of inquiry. Thus, the message that I have sought to convey in my exchange with Cohen is not that his own theory is false, but that his campaign against theories that employ cardinal probabilities to analyze factfinding in the legal process has yet to score a decisive victory. Such theories already have sufficent richness to produce insights on the law of evidence, and they are amenable to further refinement and supplementation.