Formalism and Potential Surprise: Theorizing About Standards of Proof

Dale A. Nance*

For his honorary conference, my dear friend Michael Risinger has once again articulated his distrust of "formalism" in the context of using probability theory and the associated decision theory as tools for the modeling and analysis of evidence and inference at trial. In their place, he suggests recourse to the concept of "potential surprise," and he begins to articulate how this concept could be used to explain the standards of proof to fact-finders.[1]

In my comment on Michael's suggestion, I begin by making a few comments about formalism and its relevance to Michael's thesis. Then I will turn to the concept of potential surprise and ask how recourse to this idea might advance the project of understanding and improving our practices regarding the burdens of proof.[2] I suggest that appreciation of the formal properties of potential surprise suggests lines of inquiry that can help us to decide whether it provides the basis of a superior model.

---

* Albert J. Weatherhead III & Richard W. Weatherhead Professor of Law, Case Western Reserve University School of Law.

[1] D. Michael Risinger, *Leveraging Surprise: What Standards of Proof Imply That We Want from Jurors, and What We Should Say to Them to Get It*, 48 SETON HALL L. REV. 965 (2018).

[2] In the course of my discussion, I will cite various parts of my recent book on the burdens of proof, which I cite for convenient reference here. *See* DALE A. NANCE, THE BURDENS OF PROOF: DISCRIMINATORY POWER, WEIGHT OF EVIDENCE, AND TENACITY OF BELIEF (2016).

I. THE IMPORTANCE OF FORMALISM IN THE ANALYSIS OF EVIDENCE

In an earlier article, Michael summarized his perspective on formalism in the analysis of evidence in this succinct passage:

> Formal symbolization, and its common implication of an underlying mathematizability, has great and fecund power when something approaching defensible numerical values are or can be made available. This power is what justifies the loss in general access to meaning for many people which rendering things in specialized symbolic language entails. But when such defensible numerical values are not available and are not likely to become available, then symbolization can easily become an act of mystification with very little benefit and the potential for much mischief.[3]

I readily concede the "potential for much mischief" to which Michael refers, though in truth all theory has such potential. But I think Michael understates the value of formalization in referring (albeit obliquely) only to its *computational* potential. In the present paper, to be sure, Michael qualifies this point in an important way by stating:

> While formal probability theory does have practical lessons to teach that may help to either explain or to criticize and improve the trial process in various ways, my view has always been that those lessons come by way of analogy, and that absent the availability of hard data, one must be suspicious of models that need to be supplied with cardinal numbers to work.[4]

Unfortunately, Michael does not elucidate the idea of "lessons by way of analogy." What follows may (or may not) be part of what he has in mind in this regard.

Suppose that two variables, x and y, are in a relationship of proportionality, as indicated by the standard formalization:

$$y = mx$$

for some positive constant m. Now it might seem that, in order to employ a descriptive or prescriptive model using such a relationship, one would have to have what Michael refers to as "hard data" about the values of at least two

[3] D. Michael Risinger, *Against Symbolization*, 11 LAW, PROBABILITY & RISK 247, 248 (2012).

[4] Risinger, *supra* note 1, at 968.

of these, such as m and x, in order to generate a value for the third, in this case y. But such a proportionality model can be useful even if one does not know the exact or even approximate values of the variables. For example, suppose that one can confidently say that, for two values of x, say x_1 and x_2, that $x_2 > x_1$, even though one cannot say what the values of x_1 and x_2 are. Knowing only that m is positive, the model permits certain inferences about the corresponding values of y, viz. y_1 and y_2. Specifically, one can infer that $y_2 > y_1$. Such inferences about purely ordinal relationships can be quite useful.[5]

Indeed, little more is needed to make important use of the probability-based decision-theoretic model for standards of proof at trial. By one articulation of how to apply the model to adjudication, that model requires only (1) that the law-making authorities—*not* the fact-finder—settle on some value, call it r, representing the ratio of the disutility of a false-positive decision to the disutility of a false-negative decision, and (2) that the fact-finder be able to determine whether the ratio of the probability that the claim is true to the probability that the claim is false (that is, the odds on the claim) is greater than r.[6] This does not require the fact-finder to determine a specific value for either probability, nor does it even require the fact-finder to determine a specific value for their ratio (the odds); it requires only that the fact-finder be able to say that the odds exceed r. While I agree with Michael in opposing "unwarranted cardinality,"[7] it is difficult to see how asking the fact-finder to assess whether the odds that a claim is true are greater than a certain ratio—for example, 1:1, 2:1, or 5:1—involves an *unwarranted* degree of cardinality.[8] Of course, a variety of considerations may motivate the law-making authority to use less precise verbal predicates for the value of r, which reduces the degree of cardinality further.[9] But it is not implausible to identify an r of 1:1 as grounding the "preponderance of the evidence" standard, or to identify an r of 2:1 or 3:1 as grounding the "clear and convincing evidence" standard, or to identify an r of 4:1 or 5:1 as

---

[5] Michael seems to be comfortable with the idea of ordinal rankings. *See* Risinger, *supra* note 1, at 981.

[6] To be sure, the first part of this task is rather more complicated than this description suggests, but the details do not affect the point being made here. *See* NANCE, *supra* note 2, at 21–42.

[7] Risinger, *supra* note 1, at 978.

[8] Keep in mind that the fact-finder does not set the ratio; some law-making authority does. That lawmakers are capable of doing so is no more problematic in principle, though it may be more complicated in practice, than that they can meaningfully set a certain number as the minimum age for a driver's license or the right to vote. Such numbers involve complex considerations and, once selected, are no doubt over- and under-inclusive in application to some particular cases, but that is not necessarily a fatal defect of the enterprise. *See generally* FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES (2003).

[9] *See* Nance, *supra* note 2, at 30, 40–42.

grounding the "beyond reasonable doubt" standard.[10]

In saying that such ratios do not involve unwarranted cardinality, I am primarily replying to what I take to be one of Michael's main concerns, related to the way in which fact-finders can express their degrees of certainty and whether this can or should take a quantitative form. But there are other senses in which the use of such ratios might be thought to entail unwarranted cardinality. In particular, it might be thought that, even if such a ratio can be estimated by the fact-finder, at least in the comparative terms described above, still not all of what we are concerned about in the assessment of proof can be captured by such a judgment. I actually think there is much to be said in favor of such an argument, but it does not necessarily entail abandoning such comparative assessments of probability. (Indeed, it is an objection to any theory of the standards of proof that is purely comparative, but does not necessarily entail abandoning tests of comparative assessment.) Rather, it involves recognizing that such requirements on comparative assessments of probability (or plausibility, or whatever) must be *coupled with* distinct requirements on what might be called the degree of *completeness* of the evidence presented to the fact-finder, the evidence upon which the comparative assessment is made. In other words, even if the claimant's case is sufficiently more likely than the defendant's, that may be a conclusion that is derived from evidence that is less thoroughly developed than it should be, given the nature of the case, an idea that generates a distinct component of the burden of proof. I have elaborated on this idea elsewhere.[11]

---

[10] With regard to the criminal law standard, an *r* of 4:1 is equivalent to requiring a probability of guilt of about 0.80, which may seem low to many commentators. And it may *be* low as a prescriptive matter. But as a description of jury behavior under the "beyond reasonable doubt" instruction, it may be fairly accurate, even high, at least for some categories of prosecutions. In this regard, empirical studies of mock jurors have devised ways to ascertain at what level of assessed probability a criminal conviction becomes likely. *See, e.g.*, Dale A. Nance & Scott B. Morris*, Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability*, 34 J. LEGAL STUD. 395, 407 (2005) (estimating the minimum pre-deliberation assessed probability for a stranger-rape conviction in the specific case as falling between 0.75 and 0.85); David H. Kaye, Valerie P. Hans, B. Michael Dann, Erin J. Farley & Stephanie Albertson, *Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Match Probabilities*, 4 J. EMPIRICAL LEGAL STUD. 797, 818–24 (2007) (estimating the minimum pre-deliberation-assessed probability for a robbery conviction in the specific case at 0.68).

[11] *See* NANCE, *supra* note 2, ch. 4 (arguing, inter alia, that this distinct component should be associated with the burden of production, rather than the burden of persuasion). Insofar as the use of surprise as the measure of uncertainty is also intended to incorporate considerations of evidential completeness, in the sense described in the text, it may involve what I have described as an attempt to combine into one "holistic" assessment two very different, indeed largely independent epistemic components of decision-making. If so, this can only lead to confusion. *See id.* at 156–78, 253–70. Whether this is the case will depend on how the concept of surprise is developed.

Another point is equally important. Formalism is no more inherent in probability theory and conventional decision theory than it is in the theory of potential surprise. In particular, one may approach either probability or potential surprise in informal ways, using ordinary language, verbal analogies, and so forth. Alternatively, one may approach either concept more formally, using the structured language of mathematics. Michael is certainly aware of this, for he refers to the modern formal development of the theory of potential surprise under the name "possibility theory."[12] But he makes no explicit use of that theory. For reasons that will appear below, I think this is unfortunate. In any event, it is not as if one flees the formality of probability by embracing the informality of potential surprise. There are two distinct issues here: (1) how formal (or informal) one chooses to be in analyzing the concepts to be employed; and (2) which theory–probability, potential surprise, or some other measure of partial belief—is the most useful conception in this context.[13]

With regard to the first issue, one encounters an unavoidable tension. On the one hand, if one wants to stay close to the language of the courtroom, there is reason to prefer informal understanding of terms like "probability," "possibility," and "surprise." On the other hand, formal theory is critical to understanding the strengths and weaknesses of the various alternative theories of partial belief being considered. It is an effort to articulate the intellectual commitments that lie within our informal concepts, whether that be probability, potential surprise, or any other concept. Without that effort, for example, one can make just about any argument as well with probability language as with potential surprise language, and conversely.[14] Thus, any event that has very low epistemic probability on the available evidence can be described as an event that, if discovered to be true, would cause great surprise to a rational person with that evidence, and any event of very high epistemic probability on the available evidence can be described as an event that, if discovered to be true, would cause very little surprise to a rational person with that evidence.[15] In the end, without formal theory Michael's

---

[12]   Risinger, *supra* note 1, at 981, n.49.

[13]   Michael clearly assumes that the law is focused on a gradational notion of strength of belief, not with what philosophers call "categorical" beliefs. On this point, I agree. *See* NANCE, *supra* note 2, at 278–91 (arguing that the law's norms and structures do not fit a focus on categorical beliefs).

[14]   I suspect that one reason many lawyers, surely not including Michael, are hostile to formalism is that it functions to *constrain* the arguments that can be made. Lawyers do not like that kind of rigor: they want to remain free to make any argument, no matter how incoherent, that they think nonetheless will be persuasive. What is often called "thinking like a lawyer" is not a serious impediment to such practices.

[15]   The concept of an "epistemic probability" does not make much of an appearance in Michael's paper. Like many legal scholars, he seems to see probabilities as either frequentist or subjective. Although that represents an improvement over those who see probabilities in

proposed move to potential surprise is likely to amount to little more than a change in terminology, with the attendant costs of doing so, but perhaps without an identifiable change in either the predicted or prescribed outcomes of cases.[16]

Moreover, insofar as the use of potential surprise is intended as normative—not simply an effort to conform to the extant language or thought patterns of fact-finders and other courtroom actors, but intended to improve legal decisions, at least at the margin—formal theory helps to clarify what the use of the concept is intended to achieve. Michael's proposal is not entirely clear on the matter: he seems to be suggesting that potential surprise be used not only because it captures the actual thought processes of ordinary fact-finders, but also because the rational fact-finder *should* think in these terms. Thus, he writes that "when humans evaluate evidence and determine what they believe in regard to facts, the primary, though usually implicit, operator in those determinations is, or at least ought to be, the fundamental emotion of surprise."[17] So I take Michael's enterprise to be neither wholly descriptive, nor wholly prescriptive, but rather an interpretive one. I now turn to the proposal itself.
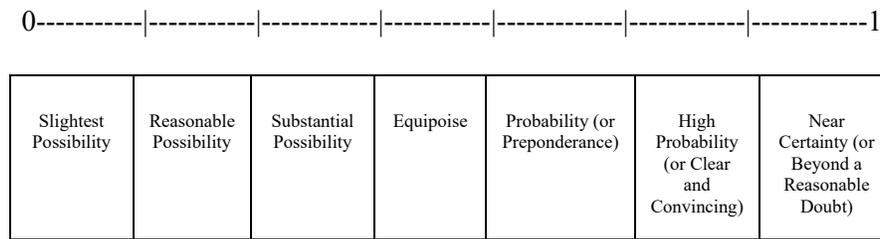
---

exclusively frequentist terms, these are still not the only alternatives, and, in my view, neither captures what is of importance in adjudication. Epistemic probabilities are degrees of rational belief or degrees of justification of belief, which are more objective than subjective probabilities, but less objective than probabilities that are identified with frequencies or propensities in observable phenomena. *See* NANCE, *supra* note 2, at 42–57.

[16] For a discussion of potential differences, see the contribution to this conference by Matthew Ginther and Edward K. Cheng. Matthew Ginther & Edward K. Cheng, S*urprise vs. Probability as a Metric for Proof*, 48 SETON HALL L. REV. 1081 (2018).

[17] Risinger, *supra* note 1, at 970.

II. POTENTIAL SURPRISE AS THE DETERMINANT UNDER THE BURDEN OF
PERSUASION

The three most commonly encountered standards of proof in litigation—"preponderance of the evidence," "clear and convincing evidence," and "beyond a reasonable doubt"—can be interpreted as representing increasing levels of probability on the right side of a scale that look something like this:[18]

0-----------|-----------|------------|-----------|-------------|------------|------------1

Slightest Possibility	Reasonable Possibility	Substantial Possibility	Equipoise	Probability (or Preponderance)	High Probability (or Clear and Convincing)	Near Certainty (or Beyond a Reasonable Doubt)

This interpretation bears an obvious relationship to the decision-theoretic model. That model provides a way to approach the allocation of the standard of proof among these tiers according to a rough weighting of the risks of error.[19]

To be sure, the diagram is not without its difficulties. Most importantly, there is a sense of "equipoise" that does not mean something like "probabilities for and against that are indistinguishably close to each other, and thus to 0.5." Equipoise should be distinguished from equi-probability. Equipoise should be understood as referring to any assessment of probability that is too close to a margin (any margin relevant under the applicable standard of proof) for the fact-finder to be able to say into which of two adjacent levels it falls. Under such a conception, there should be a zone of equipoise between each of the other six levels of probability. That is, there are five regions of equipoise, only one of which is functionally relevant under the given standard of proof. For example, if the burden of proof is "beyond a reasonable doubt," then the important cases of equipoise are those in which the fact-finder is unable to decide whether to place the case before it in the "beyond a reasonable doubt" category or in the "clear and convincing evidence" category.[20]

---

[18] *See, e.g.*, KEVIN M. CLERMONT, STANDARDS OF DECISION IN LAW 36 (2013).

[19] *See* NANCE, *supra* note 2, at 57–63.

[20] Another difficulty, significant given the present topic, is that the term "possibility" might be misunderstood: in the illustration, it should be understood as another term for degree of probability, not as the technical concept used in possibility theory, which is described below.

In any event, Michael would reject this approach. He argues that the law's references to what is "probable" do not necessarily mean probability in the formal sense—i.e., probabilities conforming to the Kolmogorov axioms of standard mathematical probability that are used in conventional decision theory.[21] Instead, he proposes to center our attention on the concept of potential surprise, and he suggests a different, more elaborate scale of levels of potential surprise to capture the levels of partial belief with which the law is concerned.[22] He explains that his "central claim is that people believe something to be true to the extent that they would be surprised to find out it was false."[23]

What motivates this proposal? Michael gives two explanations. The first relates to the goals of the law's articulation of standards of proof. He writes:

> [W]hile I accept that standards of proof are the law's attempt to impose a requirement of different degrees of belief for different kinds of issues, this does not mean that any statement about appropriate system performance over large numbers of cases was the purpose for the creation of those standards. Rather, it seems that the purpose of such standards at their creation was to define the level of subjective certainty necessary for such a decision to be a morally justified decision.[24]

Thus, Michael's rejection of the conventional decision-theoretic interpretation seems to be premised on a claim about what the purpose of the promulgation of such standards was "at their creation."

Insofar as this is correct, it seems to ignore two related points: first, a practice can continue long after its original purpose has been lost or forgotten, and for different reasons; and second, practices can be assessed and interpreted according to their functions even though those functions may not come to be understood until long after the practice is begun. The interpretation of legal doctrine often presents these phenomena. Except for the purest originalists, the origins of a doctrine are not determinative. More importantly, Michael's argument poses the following questions: On what basis would law-making authorities, then or now, think these standards of proof were or are pertinent to a "morally justified decision" if they do not speak to the relative weightings of error costs? When uncertainty about the facts is the problem, what else is morally important?   Such relative

---

[21]   *See* Risinger, *supra* note 1, at 978.
[22]   Risinger, *supra* note 1, at 985 (diagram).
[23]   Risinger, *supra* note 1, at 981.
[24]   Risinger, *supra* note 1, at 980.

weightings are, of course, built into the decision-theoretic formulation.[25] They may be quite intuitive, moral weights, or they can be the product of more thorough legislative investigation, reflection, and debate. In particular, the model does not assume (though it is compatible with) a focus on "system performance over a large number of cases."

These observations reflect a broader point. Michael has not explained how the selection of a particular segment of, or position on, the scale of potential surprise is related to the moral justification of the resulting decision. Without the decision-theoretic model, one needs an alternative way to explain, at least in a rough way, the morality of selecting some standard of proof from the available options. I understand his suggestion of using paradigmatic stories as exemplars of various levels of surprise,[26] but even if such stories can be devised, it begs the question of how moral justification is associated with the selected stories, other than by intuition.[27]

Indeed, we need some explanation about why the essentially epistemic enterprise of fact-finding in adjudication should be tied to what Michael regards as an *emotion*.[28] The objection can be made that adjudication is supposed to be, to the extent possible, a rational enterprise, not an emotional one. This is not a trivial problem, as much of the law of evidence is about channeling fact-finders toward rational and away from emotional responses. Moreover, the line between surprise and, say, anger or disgust must be difficult to draw consistently. What one fact-finder might consider surprising, another might describe as disgusting. To be sure, an emotional response can be used as a proxy for more rational epistemic measures we want adjudication to reflect, and potential surprise might be a plausible emotive proxy in this sense. In what follows, I will assume that potential surprise can be understood as such an epistemic measure or at least as a proxy for one.

---

[25] Obviously also important—and properly included in the decision-theoretic model— are positive utilities associated with true positives and true negatives. *See* NANCE, *supra* note 2, at 23–25. I put aside the important but minimal deontological constraint, deriving from the rule of law itself, that would limit the imposition of sanctions, whether civil or criminal, to those situations in which the claim is at least more likely than not true. *See* NANCE, *supra* note 2, at 33. I also put aside the possibility that controlling error costs might be subordinated to using standards of proof to send signals that control ex ante behavior. *See* NANCE, *supra* note 2, at 64–73.

[26] *See* Risinger, *supra* note 1, at 983.

[27] It should also be observed that, if stories can be used to guide jurors in matching potential surprise to categories of required proof, the same presumably would be true for probability: one could also use stories designed to illustrate the levels of probability that match categories of required proof.

[28] *See* Risinger, *supra* note 1, at 968 n.16.

Michael's second argument is quite different. It concerns how fact-finders develop their degrees of partial belief and how the law elicits those partial beliefs from fact-finders so as to resolve disputes. With regard to the former, Michael expresses a healthy skepticism about what he takes to be the Bayesian approach to a fact-finder's process of arriving at the odds on a claim, namely, the explicit use of Bayes's Theorem to revise the odds in light of the accumulating evidence.[29] I am inclined to think this is a largely a straw man, though it is certainly a recurring one. The convincing reply to this argument, available to someone who wants to insist on the usefulness of conventional decision theory, is to observe that there are many mental tools that a fact-finder will and should employ in attempting to assess a probability, or in this case, the relative probability that the legal claim is true as opposed to false (that is, the odds on the claim). Among those tools may well be gauging one's level of surprise were the claim to turn out false. Another might be the employment of an approximate probabilistic calculation using a more or less intuitive understanding of Bayesian principles. These are clearly not the only possibilities. The study of decision-making "heuristics"—more or less efficient mental shortcuts—has opened a broad field of possibilities regarding how fact-finders might go about assessing the odds on a claim.[30] Indeed, Michael elsewhere endorses an appropriate eclecticism regarding models of inference.[31]). In any event, there is nothing about the standard decision-theory model that requires the fact-finder to reach its assessment of the odds on the claim by explicitly Bayesian or any other quantitative methods.[32]

What, then, about the latter part of this argument, regarding the law's elicitation of the fact-finder's resulting partial belief about the claim, however the fact-finder may have arrived at that partial belief? As to this, Michael argues:

---

[29] Risinger, *supra* note 1, at 971–77.

[30] *See, e.g.*, GERD GIGERENZER, PETER M. TODD, & THE ABC RESEARCH GROUP, SIMPLE HEURISTICS THAT MAKE US SMART (1999).

[31] *See* Risinger, *supra* note 1, at 968 n.6. Few scholars, no matter how enamored of probabilistic approaches they may be, have seriously suggested requiring a computationally Bayesian approach by fact-finders to the assessment of every piece of evidence submitted at a trial. The most common suggestion by those endorsing the applicability of probability principles has been that, in particular contexts, such as forensic science, it can be helpful to isolate an item of evidence and invite the fact-finder to analyze it using Bayesian principles. *See, e.g.*, C.G.G. AITKEN & FRANCO TARONI, STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENTISTS (2d ed. 2004).

[32] *See, e.g.*, Kevin Clermont, *Standards of Proof Revisited*, 33 VT. L. REV. 469, 470, 482–85 (2009); Peter Tillers & Jonathan Gottfried, United States v. Copeland*, 369 F. Supp. 2d 276 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond Reasonable Doubt is Unquantifiable?*, 5 LAW, PROBABILITY & RISK 135, 142–43 (2006). *See also* NANCE, *supra* note 2, at 84–95.

> [T]he extent of surprise one would feel upon discovering a belief to be false is the best measure of subjective certainty, of the degree of belief. I propose that a person's own degree of belief is best revealed to his or her self not by the betting exercises usually invoked in Bayesian approaches such as decision theory, but by a different mental experiment—not by asking what one would bet on the truth of a belief, but asking directly how surprised one would be to find out that the thing believed was false.[33]

What gives Michael's proposal real bite is the implication that, whatever the inferential tools employed, all should be merely means to the end of gauging the fact-finder's potential surprise, rather than the fact-finder's inevitably subjective assessment of epistemic probability or odds. To the extent that this is Michael's thesis, and it appears to be, then several difficult questions emerge.

Most importantly, the notion of potential surprise is not itself without controversy and confusion. Michael draws on the work of G.L.S. Shackle, which has come to be understood as a progenitor of modern possibility theory.[34] A little possibility theory turns out to be quite useful here.[35]

An event or hypothesis is maximally potentially surprising just when it is minimally possible, and it is minimally potentially surprising just when it is maximally possible. If one takes 0 as the measure of belief in an event or proposition that the decision-maker considers minimally possible (maximally surprising), and normalize (set to the value of 1) the measure of belief in an event or proposition that is maximally possible (minimally surprising), then possibility measures vary from 0 to 1 with *pos(A) = 0* interpreted as "*A* is impossible" (maximally surprising) while *pos(A) = 1* means that "*A* is perfectly possible" (not surprising at all). In considering a set of mutually exclusive and exhaustive hypotheses, complete ignorance is when all the hypotheses are completely possible, whereas complete knowledge is when all but one of the hypotheses are impossible (in which case the one hypothesis must be perfectly possible).

Without getting too technical, the central premise of possibility theory can be identified this way. If *A* is a collection of hypotheses or propositions (such as stories supporting a litigant), then possibility is that concept of partial belief that selects the most believable proposition from among those in *A* and attributes the same degree of belief to *A* that is attributed to the most believable proposition in *A*. Thus, if *A* is some finite collection of stories,

---

[33]   *See* Risinger, *supra* note 1, at 981.
[34]   *See* Risinger, *supra* note 1, at 971–73.
[35]   For those interested, there is an excellent collection of papers on this and related subjects. *See* DEGREES OF BELIEF (Franz Huber & Christoph Schmidt-Petri eds., 2009).

each of which is assigned a possibility value, then
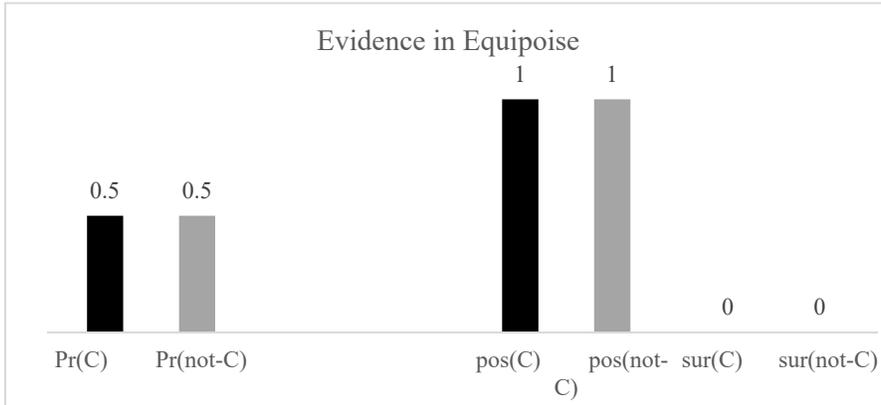
$$pos(A) = \max_{\alpha \, \varepsilon \, A} pos(\alpha)$$

When the possibility of *A* attains the maximum possible value (perfectly possible, normalized to 1), it does not entail that the possibility of *not-A* is minimized, i.e., 0. That is, it is not necessarily the case that *pos(A) + pos(not-A) = 1.*[36] This is a central distinction between probability and possibility. Possibility theory is about belief measures that are "non-additive" in this sense. As a result, to adequately describe a person's epistemic situation under possibility theory, one needs to know two things: not just the person's assessment of the possibility that a proposition is true, but also the person's assessment of the possibility that the proposition is false. The latter may not simply be derived from the former by a direct calculation. Again, the same point applies to potential surprise: a full description of a person's doxastic state with reference to an event must include both the potential surprise if the event is true and the potential surprise if the event is false.[37]

These abstractions may be somewhat more digestible by way of an illustration of the different dynamics of probability and possibility/surprise measures. The first graphic that follows compares the situation of "equipoise" in a civil case and how this might be reflected in the strength of the probability measures (denoted *(Pr(·))*) and possibility/surprise measures for the claim, *C*, and against the claim (for the negation of the claim), *not-C*:
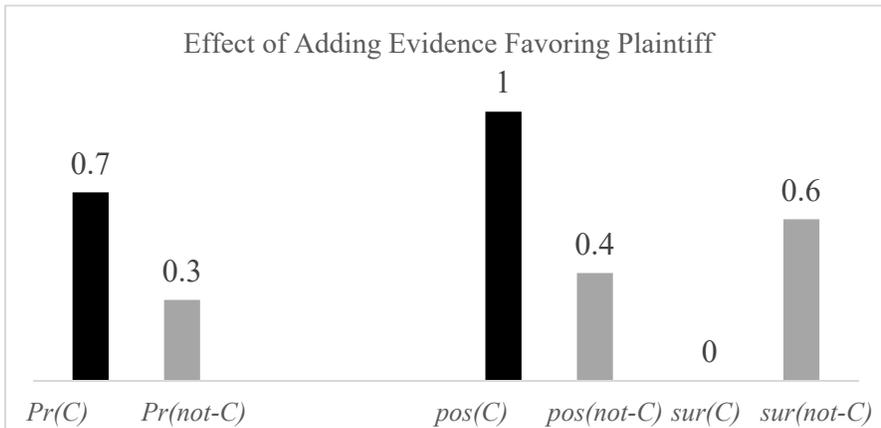
---

[36] Indeed, for any *A*, if *pos(A) < 1*, then *pos(not-A) = 1*. This is because the most plausible event, normalized to have a possibility of 1, must lie somewhere; if not in *A*, then it must be in *not-A*. (Some scholars use an axiom system that would require an additional condition on this implication, namely, that *pos(Ø) < 1*. This condition excludes the possibility that the most plausible event had not been contemplated and, thus, lay in neither *A* nor *not-A*, but rather in Ø. *See* Donald Katzner, *Potential Surprise, Potential Confirmation, and Probability*, 9 J. POST KEYNESIAN ECON. 58 (1986).

[37] If *sur(A)* represents the surprise function value for a collection of stories, *A*, then *sur(A) = 1 – pos(A)*. Accordingly:

$$sur(A) = \min_{\alpha \, \varepsilon \, A} sur(\alpha)$$

Evidence in Equipoise

	Pr(C)	Pr(not-C)		pos(C)	pos(not-C)	sur(C)	sur(not-C)
	0.5	0.5		1	1	0	0

Now imagine that evidence favoring the claim is introduced. The effect under a probability measure will be to increase the probability of *C* and to decrease the probability of *not-C*, as indicated on the left of the second graphic. (The numbers used are merely illustrative, though they must sum to 1.0.) But using possibility measures, the effect will not increase the possibility of *C*; this measure is already as high as it may go. Rather, the effect of such evidence is to drive down the possibility of *not-C*. (Put in terms of surprise, such evidence cannot decrease *sur(C)*; it may only increase *sur(not-C)*.)[38]

Effect of Adding Evidence Favoring Plaintiff

	Pr(C)	Pr(not-C)		pos(C)	pos(not-C)	sur(C)	sur(not-C)
	0.7	0.3		1	0.4	0	0.6

---

[38] Again, the numbers used are merely illustrative. In particular, I am not asserting anything about the quantitative relationship between changes in probability and changes in possibility measures. However, it is the case that *sur(not-C) = 1 –pos(not-C)*.

So how does this relate to Michael's proposal? Although Michael does not talk about the two-parameter feature of possibility theory (or potential surprise theory), it seems to be embedded in his proposal. As above, let *C* (for "claim") represent what Michael calls the "factual proposition or complex of propositions" at issue—for example, the proposition that the defendant committed the alleged crime. (Think of this as the collection of all stories that, if true, would instantiate the offense with which the defendant is charged.) Then, inspection of Michael's proposed scale of potential surprise reveals that the fact-finder's potential surprise at *C* being true is the focus of one half of the spectrum, with the extreme points being "absolute falsity" of *C* on one end and equipoise on the other. But on the other half of the spectrum, ending at the point "absolute truth" of *C*, the focus shifts: it seems, instead, to be concerned with the potential surprise at discovering *not-C*. Put in the language of possibility theory, the former end of the scale measures *pos(C)*, with an extreme point where *C* is impossible (*pos(C) = 0*), while the latter end of the scale measures *pos(not-C)*, with an extreme point where *not-C* is impossible (*pos(not-C) = 0*).

This invites two questions. What happens to *pos(C)* when one moves into the latter portion of the scale? And conversely, what happens to *pos(not-C)* when one moves into the former portion of the scale? A possible answer to what Michael has in mind comes from possibility theory: *pos(C)* reaches its maximum (normalized to 1) at what Michael calls equipoise and remains at 1 for the rest of the scale. Similarly, *pos(not-C)* increases as one moves in the opposite direction on the scale, reaching its maximum of 1 at equipoise and remaining there. Thus, Michael's scale seems to be two possibility scales superimposed. Here is an abbreviated form of the two possibility scales I think are embedded in Michael's single scale, reducing the number of divisions on each scale (from 15 to 7) for simplicity and to facilitate comparison with the probability scale presented earlier:

0------------|-------------|-------------1---1

Slightest Possibility of *C*	Reasonable Possibility of *C*	Substantial Possibility of *C*	Perfect Possibility of *C*	Perfect Possibility of *C*	Perfect Possibility of *C*	Perfect Possibility of *C*

1--1-------------|--------------|-----------0

Perfect Possibility of *not-C*	Perfect Possibility of *not-C*	Perfect Possibility of *not-C*	Perfect Possibility of *not-C*	Substantial Possibility of *not-C*	Reasonable Possibility of *not-C*	Slightest Possibility of *not-C*

The first scale obviously represents, from left to right, an increasing *pos(C)*; the second scale represents, from right to left, an increasing *pos(not-C)*. Superimposed, these scales yield the following:
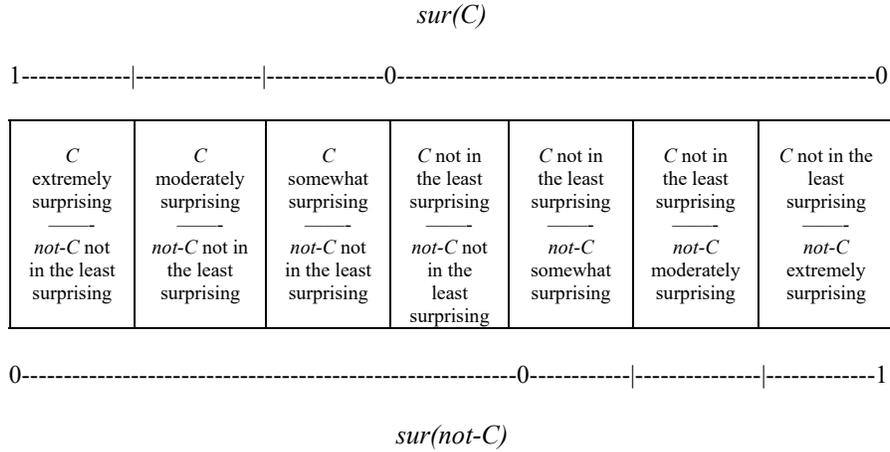
pos(C)

0------------|--------------|-------------1---1

Slightest Possibility of *C*	Reasonable Possibility of *C*	Substantial Possibility of *C*	Perfect Possibility of *C*	Perfect Possibility of *C*	Perfect Possibility of *C*	Perfect Possibility of *C*
Perfect Possibility of *not-C*	Perfect Possibility of *not-C*	Perfect Possibility of *not-C*	Perfect Possibility of *not-C*	Substantial Possibility of *not-C*	Reasonable Possibility of *not-C*	Slightest Possibility of *not-C*

1--1-------------|--------------|-----------0

pos(not-C)

Here is the same scale roughly translated into potential surprise, denoted as *sur(C)* and *sur(not-C)*:

sur(C)

1------------|--------------|-------------0--0

C extremely surprising ——- *not-C* not in the least surprising	*C* moderately surprising ——- *not-C* not in the least surprising	*C* somewhat surprising ——- *not-C* not in the least surprising	*C* not in the least surprising ——- *not-C* not in the least surprising	*C* not in the least surprising ——- *not-C* somewhat surprising	*C* not in the least surprising ——- *not-C* moderately surprising	*C* not in the least surprising ——- *not-C* extremely surprising

0---0-----------|--------------|------------1

sur(not-C)

Perhaps this is not where Michael is headed, even if we refine the scale to 15 distinct levels, as Michael has it, rather than the simpler 7 illustrated here. He does not emphasize the shift in focus from *sur(C)* to *sur(not-C)*, and the labeling of the segments of his proposed scale does suggest that there is only a single epistemic measure involved, one that is steadily increasing (or decreasing) from one end of the scale to the other. This is not the case in my version. But if I am correct about what Michael is trying to do, or if it is what he would endorse upon further reflection, there are several things to notice.

This picture makes sense in Shackle's way of looking at things. At any point in the process of assessing evidence, the fact-finder whose partial beliefs conform to possibility theory must have a doxastic state that can be represented on this scale. In particular, the scales of *sur(C)* and *sur(not-C)* do not need to be "orthogonal" (like Cartesian coordinates) rather than "superimposed" in order to reflect all the conceivable possibilities: it cannot be that both *sur(C)* and *sur(not-C)* are greater than 0, so those pairs of potential values need not be represented.[39]

In the pivotal central cell, both *C* and *not-C* are perfectly possible, not in the least surprising. According to Shackle, this is entirely coherent. His view was that an outcome is perfectly possible when there is no discernible obstacle to it being or becoming true.[40] Here is an example of the kind of situation he had in mind:

---

[39]   *See supra* note 36.
[40]   A perfectly possible hypothesis is "entirely unobstructed, wholly free . . . of any threatened interference." G.L.S. Shackle, *The Bounds of Unknowledge*, *in* BUSINESS, TIME AND THOUGHT 86 (Stephen F. Frowen ed., 1988).

> An opaque urn contains 100 balls, each of which may be black or white; you have and can acquire no further information about the color of the balls until after you draw one. What do you believe about the proposition that the ball you draw from the urn will be white?

This is an example in which one's belief about the drawn ball being white is based on so little information that it is hard to form any partial belief on whether the ball to be drawn will be white. And the same is true of a partial belief that the ball will be black. Shackle described the situation as one in which the potential surprise associated with either outcome is 0, or in the language of possibility theory, *pos(white) = pos(black) = 1*. That is, both drawing a white ball and drawing a black ball are perfectly possible. Of course, if a party to a lawsuit were burdened to prove that the ball to be drawn is white, understandably there would be a failure of proof, and the burdened party should lose. This is one way to understand "equipoise," and sure enough, Michael labels this central portion of the scale "functional equipoise."

But the meaning of "perfectly possible" is not always so obvious. Consider the following example:

> An opaque urn contains 100 balls, 80 white and 20 black; you have and can acquire no further information about the color of the balls until after you draw one. What would you believe about the proposition that the ball you draw from the urn will be white?

Under these conditions, is it perfectly possible that the ball to be drawn will be white? Is it perfectly possible that the ball to be drawn will be black? Scholars familiar with Shackle's theory debate which of the following accounts is correct:

- By one account, there is no discernible obstacle to obtaining a white ball and there is no discernible obstacle to obtaining a black one, so each outcome is perfectly possible.[41] (This would place the situation in functional equipoise.)

- By another account, the fact that one outcome is more probable than another implies that there is a discernible

---

    [41]   *Cf.* Jochen Runde, *Shackle on Probability*, in ECONOMICS AS AN ART OF THOUGHT: ESSAYS IN MEMORY OF G.L.S. SHACKLE 226 (Peter E. Earl & Stephen F. Frowen eds., 2000) ("Compare the toss of a fair coin with the toss of a fair die. The outcomes of a head and an ace are both perfectly possible on Shackle's definition, although the respective probabilities are different.").

obstacle in the way of obtaining the latter, so the former must be assigned a higher possibility, and the latter cannot be perfectly possible.[42] (This would remove the situation from functional equipoise.)

- By a third account, the presence of black balls is a discernible obstacle to drawing a white ball, just as the presence of white balls is a discernible obstacle to drawing a black ball, so that neither is perfectly possible, though drawing a white ball has a higher possibility than drawing a black one.[43] (This would also remove the situation from functional equipoise.)

To be sure, this is not the kind of decision context for which possibility theory was designed, because probability-based decisions are easily applied. But it is disconcerting that no clearly correct answer to such a simple problem is available under possibility theory or the theory of potential surprise.[44]

If clarification of this idea of complete possibility (or, the complete absence of surprise) is to be obtained, I suspect it will be found in fuller elaboration of the notion that this occurs when "no obstacle" to the proposition being true is discernible. Professor Ho puts the matter this way: "By 'potential surprise', Shackle means the surprise which runs *counter* to our expectation. This is different from the surprise that is felt upon encounter of *novelty*, or the *unexpected*."[45] In other words, to run counter to expectations, and thus to produce surprise, one must *have* expectations about something being true or happening. In the absence of expectations, there is zero surprise and maximum possibility. And where do expectations come from? Answer: experience with how things generally or normally work. This could be connected, then, to one of numerous presumptive reasoning theories that use patterns of normalcy, subject to defeasing conditions, to characterize

---

[42] *See* J.L. FORD, CHOICE, EXPECTATION AND UNCERTAINTY: AN APPRAISAL OF G.L.S. SHACKLE'S THEORY 72 (1983) ("To hold that the one outcome is more likely than the other must imply that there is an obstacle in the way of the other outcome's occurring. Therefore, it must be assigned a higher degree of potential surprise than the one outcome.").

[43] *See* H.L. HO, A PHILOSOPHY OF EVIDENCE LAW: JUSTICE IN THE SEARCH FOR TRUTH 147 (2008). This account violates the usual axioms of possibility theory, for it cannot simultaneously be the case that neither *C* nor *not-C* is perfectly possible, unless perhaps the most plausible story has not been contemplated as a possibility, so that it falls within neither *C* nor *not-C*. *See supra* note 36. Still, Professor Ho's argument is an intelligible reading of Shackle's criterion of perfect possibility, which reveals a tension that is inherent in the theory.

[44] Of course, the same may be said for probability-based decisions in law. The balls in an urn problem is just another metaphor for so-called "naked statistical evidence" cases. *See* HO, *supra* note 43, at 135–43.

[45] *Id.* at 147 (emphasis in original).

epistemic justification.[46] (For example: normally, in a certain situation (i.e., given certain evidence), one is warranted in believing p, unless d_1, d_2, or d_3; if, for example, d_1, then one is not normally warranted in believing p, unless d_1^a or d_1^b; and so forth.) The problem with such theories is that they end up as theories of categorical belief, which Michael has so far not endorsed as controlling in adjudication and which pose serious problems once one comes to grips with the need to synthesize partial beliefs about the presence of defeasing conditions into an overall gradational measure of epistemic warrant.[47]

On the other hand, the model suggested above has some promising features. First, the fact that there are two variables of interest at least starts us on a path that might ultimately present a solution to what Michael recognizes as a vexing problem: "how to get some sort of relative weight out of what starts out as an ordinal system."[48] The resulting scheme would bear a strong resemblance to the theory advanced by Hoc Lai Ho in his 2008 monograph. Discussing Shackle's idea of potential surprise, Ho concludes as follows:

> To believe that p, we must not only think that p is perfectly possible, we must also think that none of its contradictories is perfectly possible. How strongly one believes that p in that situation will depend on the possibility of p relative to the possibility of the strongest of the contradictories.[49]

Moving from the left to the right on my superimposed scales involves just this two-part test: if the strength of the evidence causes the fact-finder's doxastic state to reach equipoise, one is assured that the claim, C, is perfectly possible. If the strength of the evidence moves the doxastic state beyond equipoise, it at least becomes meaningful to talk in terms of an increasing ratio of *pos(C)/pos(not-C)*, and thus potentially to talk in terms of such a ratio exceeding a target value r, where $r \geq 1$.[50] Obviously, more needs to be done here to complete such a theory, but there is potential.

---

[46]   *See, e.g.*, Martin Smith, *What Else Justification Could Be*, 44 NOÛS 10 (2010).

[47]   *See* NANCE, *supra* note 2, at 137–47. It might be relatively unproblematic if this difficulty could be cabined off, confronted only in the context of delimiting the category of equipoise, but it is not easy to see how this can be done, given that every point on the spectrum is characterized by at least one of the variables having the value of "perfectly possible." Moreover, outside of equipoise, one of the variables is less than perfectly possible, and a degree of possibility must be associated with a degree of obstacles to truth.

[48]   Risinger, *supra* note 1, at 981 n.49.

[49]   HO, *supra* note 43, at 148. Professor Ho invokes Shackle to support his categorical belief model of adjudicative fact-finding. As already indicated, I do not think a categorical belief model works. *See supra* note 13.

[50]   It should be obvious that I make no claim to the effect that a fact-finder's doxastic state must move continuously along this scale, even if possibility theory effectively models

Possibility theory also articulates another idea, the "necessity" of an event. As explained above, the possibility of the truth lying within a collection of stories, *A*, is identified with the most plausible story in the collection. This means that the possibility of *not-A* is identified with the most plausible story not within *A*. This maximizing measure of *not-A* in turn permits a measure of the certainty or necessity of *A*, as the complement of the possibility of *not-A*. That is, *nec(A) = 1 – pos(not-A)*. If one defines the collection of stories satisfying the claim, *C*, as those satisfying both a proposition *A* and a proposition *B*, then one striking fact is that this measure has the following property:

$$nec(C) = nec(A \text{ and } B) = min \; [nec(A), nec(B)]$$

If one thinks of *A* and *B* as the elements of the claim, *C*, then this property provides a potential solution to the so-called conjunction paradox, or problem of the conjunction, that is thought by some (not including me) to be a serious embarrassment to thinking in terms of probabilistic standards of proof.[51] If the standards of proof are set in terms of necessity measures, then whether that standard has been met as to the claim, *C*, can be determined entirely by assessing whether it has been met as to each of *A* and *B*. Unlike a probability measure, *nes(C)* can never be less than the smaller of *nes(A)* and *nes(B)*. Necessity thus has a structure like the idea of inductive support that Jonathan Cohen proposed some time ago in an effort to solve this (and other) supposed paradoxes.[52]

This works rather nicely with the superimposed scale described above. On the left-hand side of the scale, one need not worry about *nec(C)* because the obvious obstacle preventing a verdict favoring the claimant (the party asserting *C*) is the fact that *C* is not perfectly possible. But once this condition is satisfied, attention shifts to the right-hand side of the scale, where the epistemic action focuses on *sur(not-C)*. And the surprise function has the same property described above for necessity, for the simple reason that *nec(A) = 1 – pos(not-A) = sur(not-A)*. Consequently, in deciding the second part of the two-part test for a verdict, one can use the property:

$$sur(not\text{-}C) = nec(C) = nec(A \text{ and } B) = min \; [nec(A), nec(B)]$$
$$= min \; [sur(not\text{-}A), sur(not\text{-}B)]$$

---

that fact-finder's partial beliefs. I am simply using movement along the scale as a way of explaining what the scale represents.

[51] For extended recent discussions of the problem compatible with a probabilistic understanding of the standards of proof, see David S. Schwartz & Elliott R. Sober, *The Conjunction Problem and the Logic of Jury Findings*, 59 WM. & MARY L. REV. 619 (2017); Mark Spottswood, *Unraveling the Conjunction Paradox*, 15 LAW, PROBABILITY & RISK 259 (2016).

[52] *See* L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE (1977).

Verbally, if a verdict favoring *C* is appropriate only if *not-C* would be sufficiently surprising by a designated standard, and if *C* consists of multiple elements, and if the fact-finder can identify the element that would be least surprising if false, then, in order to give a positive verdict on the claim, the fact-finder need only assure itself that that element would be sufficiently surprising if false under the designated standard.

CONCLUSION

I have gone far enough in suggesting both the problems and the intriguing lines of inquiry surrounding possibility/surprise measures as tools for interpreting legal standards of proof and what we ask of fact-finders in their application. My main point has been to suggest how formal theory can be useful to someone like Michael in his effort to articulate a better way to understand adjudicative fact-finding. Put simply, my message to Michael is this: formalism can be your friend, too. I have found it very instructive to pursue Michael's suggestion by taking a few steps toward the articulation of a coherent conception of the standards of proof using possibility theory. Of course, Michael may not choose to go this direction, but he will need to choose *a* direction for the development of his idea.

My initial conclusion is that the use of surprise as a measure of uncertainty pertinent to adjudicative fact-finding faces several obstacles. The first is descriptive: do fact-finders think in terms of potential surprise? It is rather difficult to see a confirmation of Shackle's idea in the interesting empirical data that Matt Ginther and Ed Cheng provide for us at this Symposium. My *sur(C)* is their *surprise(guilty)*, while my *sur(not-C)* is their *surprise(innocent)*. In the responses of subjects, we do not see clearly various properties that we would expect to see from Shackle-style surprise measures. First, as one moves from the weakest evidence for the claim to equipoise, *surprise(innocent)* should be 0 throughout the range, while *surprise(guilty)*, always higher than *surprise(innocent)*, should gradually fall to 0. Second, at equipoise, both *surprise (innocent)* and *surprise(guilty)* should be essentially 0. Third, as the evidence becomes yet more favorable to the prosecution, *surprise(guilty)* should remain 0 throughout the range past equipoise, while *surprise(innocent)* should rise gradually over the range from 0 toward 1. Rather than these patterns, Ginther and Cheng's data suggest that, without tutelage, respondents are inclined to understand surprise as something more like the complement of probability than like the complement of possibility, at least when a quantitative response scale is employed.[53]

---

[53]   *See* Ginther & Cheng, *supra* note 16, at 1087–90 figs. 1–4 (showing data plots).

The second set of obstacles is prescriptive. If ordinary people serving as fact-finders use surprise as the complement of probability, it is difficult to see that much is gained from switching terminology from one to the other, and there will certainly be transaction costs associated with such a change. If, instead, we want fact-finders to think in terms of Shackle-style surprise, then a real change might result, but an educative process would be required, and it is even harder to see how that could be worth the costs of the effort. Moreover, I have already noted the problems associated with identifying levels of moral justification with particular levels of potential surprise. That project must somehow explain the connection between levels of required proof and the costs of errors, both false positives and false negatives, beyond merely the observation that setting the level of acceptable risk of such errors is the (perhaps unintended) consequence of a moral justification that is developed without regard to them.

Finally, both the descriptive and prescriptive aspects of Michael's proposal require at least some effort to think through the analytical problems associated with surprise measures. In particular, if Shackle's model or the subsequently and more elaborately developed possibility theory is to form the basis of Michael's suggestion, greater clarity is needed about the logic of surprise. Especially important is the central idea of "completely possible" (or "completely unsurprising") events, without which the important idea of equipoise is left rather shrouded in mystery.

One final thought: I am immensely grateful to Michael for his consistent creativity and insight over the years, not to mention his collegiality. I trust his retirement from teaching will not mean an end to his contributions to the field of evidence. His work always inspires—and usually convinces. I hope to see much more.