What Standards of Proof Imply We Want from Jurors, and What We Should Say to Them to Get It

D. Michael Risinger*

> Tell me that the great Mr. Gladstone, in his last hours, was haunted by the ghost of Parnell, and I will be agnostic about it. But tell me that Mr. Gladstone when first presented to Queen Victoria, wore his hat in her drawing room, slapped her on the back, and offered her a cigar, and I am not agnostic at all. That is not impossible; it's only incredible. But I'm much more certain it didn't happen than that Parnell's ghost didn't appear.—G.K. Chesterton[1]

> The most exciting phrase to hear in science, the one that heralds new discoveries, isn't "Eureka!" but rather "Hmm ...that's funny."—Isaac Asimov[2]

Introduction

When Professor Laudan contacted me in the fall and asked if I would be interested in participating in what I now refer to, somewhat tongue in cheek, as the first Larry Laudan Lecture Series, I was both flattered and daunted. The "flattered" part is self explanatory, but the "daunted" part requires some explanation. I had already bitten off more than I was sure I could chew in the fall and winter, with deadlines on a book chapter and two symposia all looming before February 1. However, Larry assured me that it would not be inappropriate for me to use parts of those projects for this presentation. Although all of those pieces got finished, their appropriateness for present purposes didn't pan out, so I have taken the liberty of stretching that authorization a bit further, and I have used a short piece recently published in the journal Law, Probability & Risk, in order to put the concluding (and new) part of the this presentation into context, or the context of my current thinking on the law of proof, at any rate.[3] Hence the odd structure, with two titles (and two sets of epigrams).

---

* John J. Gibbons Professor of Law, Seton Hall University School of Law. A version of this paper was presented at the International Conference on Artificial Intelligence and Law at the University of Pittsburgh, June 10, 2011.

[1] G.K. Chesterton (speaking through Father Brown), "The Curse of the Golden Cross," in G.K. CHESTERTON, THE INCREDULITY OF FATHER BROWN 122 (1926)

[2] This is the usual attribution, but there is no printed source for it, and it well may be wrong.

[3] 11 L. Prob. & Risk 247 (2012). To that end, I have freely supplemented the footnote material (though not the text) in that article, so that this version may be thought of as a second edition, an opportunity one rarely gets with articles (or monographs either, for that matter. I would have said "books," but there is no end to the making of new editions of [case]books).

The newly created part at the end of this writing will attempt to make some contribution to the long-running debate over what the law intends by its trial standards of proof, and how best we might communicate that to a jury. In this part I will combine three elements not usually deployed for these purposes: rank ordering, the story model, and the notion of surprise.

Against Symbolization

When you can measure what you are
speaking about, and express it in numbers,
you know something about it....
—Lord Kelvin[4]

There are few, if any, useful ideas in economics
that cannot be expressed in clear English.
—John Kenneth Galbraith[5]

Liability depends upon…whether B [is] less than PL.
—Learned Hand[6]

---

[4] Sir William Thomson (later Baron Kelvin), Lecture: "On Electrical Units of Measurement" (May 3, 1883), in SIR WILLIAM THOMSON, POPULAR LECTURES AND ADDRESSES (2nd ed. 1891), V. I, p. 80.

[5] JOHN KENNETH GALBRAITH, THE NEW INDUSTRIAL STATE 419 (3rd ed. 1978).

[6] U.S. v. Carroll Towing Co., 159 F. 2d 169 (2nd Cir. 1947) (Hand, J). The "Hand formula" for negligence is a legal icon. I can't say I know why. The Carroll Towing case involved a claim that the owners of a barge that had broken away and caused damage had been negligent in allowing the barge to break away. The original quotation from Judge Hand's opinion is:

> the owner's duty, as in other similar situations, to provide against resulting injuries is a function of three variables: (1) The probability that she will break away; (2) the gravity of the resulting injury, if she does; (3) the burden of adequate precautions. Possibly it serves to bring this notion into relief to state it in algebraic terms: if the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P: i.e., whether B less than PL.

What of value the symbolization adds to the ordinary language explanation is by no means clear. What is clear, at least to me, is that the act of generating the formula covered up the assumptions behind the assertion of a simple multiplication relationship, and the assumption that definitions of the variables might yield empirically warranted numbers that would allow such a simple multiplication. See the similar problem discussed *infra* at note 19. At any rate, the Hand Formula became the darling of the Law and Economics movement, which quickly rendered it in fully symbolized glory thus: "an act is in breach of the duty of care if: $B < PL$." *See* Calculus of Negligence, http://en.wikipedia.org/wiki/Calculus_of_negligence.

When Jeremy Bentham wanted to summarize his felicific calculus, he wrote a mnemonic verse.[7] When modern rational choice and expected utility theorists do the same, they write a formally symbolized expression. I want to suggest that Bentham's instinct in this regard was superior. 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.[8]

A lot of what I wanted to say about the limitations of an artificial intelligence (AI)[9] program directed toward evidence and inference (and their use in legal proceedings) has already been said by Ron Allen and Jim Franklin in their submissions for this conference.[10] So perhaps the best course is simply to put forward a series of propositions as theses and see where they

---

[7] Intense, long, certain, speedy, fruitful, pure—
Such marks in pleasures and in pains endure.
Such pleasures seek if private be thy end:
If it be public, wide let them extend
Such pains avoid, whichever be thy view:
If pains must come, let them extend to few.

JEREMY BENTHAM, THE PRINCIPLES OF MORALS AND LEGISLATION (1823 edition) (with author's final additions and emendations), p. 27, fn. 1.

[8] "It is important… to keep in mind that the use of mathematical language is not a goal for its own sake in the natural sciences, but is a highly specialized technique for arriving at a very complete consensus. Any attempt to apply the same intellectual technique to social and behavioural phenomena must have the same purpose, and must therefore maintain comparable criteria of categorical precision and operational fidelity to the realities. By these criteria, unfortunately, most such attempts appear implausible and pretentious, and probably induce obfuscation and mental opacity where intended to clarify and persuade."
JOHN ZIMAN, RELIABLE KNOWLEDGE: AN EXPLORATION OF THE GROUNDS FOR BELIEF IN SCIENCE 166 (1977).

[9] One might charge me with adopting unnecessary formal symbolization when I use the initialism "AI." However, acronyms or initialisms are the most trivial of symbolizations, and the nearest to ordinary language. See the discussion in note 19, *infra*. Indeed, some acronyms (or, if you prefer, initialisms) become so common with use as to become ordinary words in themselves (FBI, for instance). Such initialisms newly created for the purposes of a single piece of expository writing can become so confusing that they do not justify whatever efficiency is gained. In the present case, however, "AI" is already so common as to present few if any such problems.

[10] See James Franklin, 11 L., Prob & Risk 225, and Ronald J. Allen (in press). It might not be amiss for these proceedings to set out Professor Franklin's abstract:

Fifty years of effort in artificial intelligence (AI) and the formalization of legal reasoning have produced both successes and failures. Considerable success in organizing and displaying evidence and its interrelationships has been accompanied by failure to achieve the original ambition of AI as applied to law: fully automated legal decision-making. The obstacles to formalizing legal reasoning have proved to be the same ones that make the formalization of commonsense reasoning so difficult, and are most evident where legal reasoning has to meld with the vast web of ordinary human knowledge of the world. Underlying many of the problems is the mismatch between the discreteness of symbol manipulation and the continuous nature of imprecise natural language, of degrees of similarity and analogy, and of probabilities.

lead. When Luther (figuratively) nailed his 95 theses to the Church's door (if not to a church door[11]), he was committed to their truth. I am much more tentative about what follows. These propositions are put forward rather loosely, and will often be accompanied by small expansions or illustrations. Many of these propositions will be generalities that are quite contestable in detail, but I hope that they might add up to something of interest. Even if not, critical responses to them will be of great interest and educational value to me. They are talking points not so much for me to talk about to inform others, but for others to talk about to inform me.

---

1. Natural language is not yet describable or modelable as a formal system, by which I mean an axiomatized symbolic system with decision rules that determine true and false statements (i.e. symbol strings that are or are not theorems) within the system.[12]

2. Computability is dependent on such a formal modeling at some level.[13]

3. The important aspects of natural language that give it its power may never be describable as a formal system. One of these is the inherent property of words that reflects meaning sufficiently for most communication, but is sufficiently indeterminate to allow for and

---

[11] It appears doubtful that any such dramatic nailing to an actual door occurred. *See* http://en.wikipedia.org/wiki/The_Ninety-Five_Theses

[12] This is, I believe, a fairly standard account of the notion of a fully formalized system. *See* DOUGLAS R. HOFSTADTER, GÖDEL, ESCHER, BACH: AN ETERNAL GOLDEN BRAID 33–60 (1979). As to some of the difficulties in computer modeling natural language thus far, *see* Franklin, *supra* note 10. *See also* STEVEN PINKER, WORDS AND RULES: THE INGREDIENTS OF LANGUAGE 83–119 (1999). For a summary of the current status of computer handling of natural languages in such very useful but low-order tasks as translation and text classification, *see* Stuart Russell and Peter Norvig, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 860–85 (3rd ed. 2010).

[13] All current common approaches to computing involve reductions to formal binary expressions (there are only very minor research efforts exploring analogue computing. *See* "Analogue Computer," http://en.wikipedia.org/wiki/Analog_computer). Thus even multi-valued, continuous and "fuzzy" relationships have to be simulated in binary to be computable. Apparently this remains true even in the realm of quantum computing, where a single gate can exist in something other than a binary state, so that in spite of this the only advantage of a quantum Turing machine over a conventional Turing machine is realizable speed. At least, this is how I now understand the current state of play. *See generally* the discussion of the Church-Turing thesis on computability, the Turing machine and the universal Turing machine in JOHN R. SEARLE, MIND: A BRIEF INTRODUCTION 68–70 (2004), and the quantum Turing machine in Quantum Computer, http://en.wikipedia.org/wiki/Quantum_computer. Of course, such digital modeling of continuous processes is perfectly adequate for a wide range of applications, as anyone's CD or DVD collection well illustrates. But whether it is or will be adequate to capture human cognition is certainly an open question. One does not have to go as far as Roger Penrose in invoking the foundational nature of quantum processes, *see* ROGER PENROSE, THE EMPEROR'S NEW MIND 400–14 (1989), to have the sneaking suspicion that purely digital foundations may be inadequate. Steven Pinker considers a central strength of natural language to be the fact that it contains both a digital module and an analogue module, which helps both with analysis and with proper navigation of an essentially analogue world. *See* PINKER, *supra* note 9, at 269–87 (Ch. 10—"A Digital Mind in an Analogue World"). See also the suggestion in note 12 *infra*.

promote, among other things, fecund pattern perception, and analogical perception and reasoning, broadly conceived.[14]

4. Even if the third proposition turns out not to be true in theory (and there are perhaps good reasons to hold this view[15]), current modeling, while impressive, is in some senses profoundly trivial. Deep Blue can play chess, but not Jeopardy. Watson can play Jeopardy but not chess. Go remains intractable. A formal program that modeled natural language, and the reasoning that underlies it, would not only have to be able to play all three (although perhaps none well), but more importantly would have to be able to discuss the advantages and disadvantages of natural language over more determinate formal systems lacking the properties of natural language, spontaneously on its own.[16]

5. Current modeling resulting in programs combined with current hardware forms a package which is clearly superior to natural language, or to formal systems manipulated by humans alone, in regard to the solution to some problems that have important practical applications.[17] But to the extent that these applications actually "make decisions" that change exterior conditions in the world, the decision criteria generally must be relatively simple and fairly generally agreed upon by humans.

6. This is because such world-altering decisions are fundamentally value judgments about what is better to happen in human terms. Those value judgments are sometimes clear (shut down the nuclear reactor before it melts and kills thousands), but often not (execute the prisoner, do the abortion). Only humans ought to make such decisions for humans (even if other humans might prefer the decision resulting from the non-human program, which in actuality would be nothing more than the judgment, intended or unintended, resulting from the choices of the programmers).[18]

---

[14] The centrality of analogy and metaphor to human cognitive and linguistic processes is well explained and defended in, inter alia, GEORGE LAKOFF AND MARK JOHNSON, PHILOSOPHY IN THE FLESH: THE EMBODIED MIND AND ITS CHALLENGE TO WESTERN THOUGHT 3–73 (1999).

[15] For one such position, *see* DANIEL DENNETT, CONSCIOUSNESS EXPLAINED, pp. 210 et seq. (1991). However, I suspect that modeling human cognition, if it is ever possible, will require more reliance on analogue processing than is currently being pursued. I say this because the chemical concentrations that define the conditions of neural firing seem more easily understood as underlying analogue processes than purely digital processes and, to me at least, it seems easier to conceive of the Selfridge Pandemonium Theory processes (which seem fairly well established as at least the deep-level architecture of human processing) as being produced by such continuous phenomena. For a good summary of the current evidence of pandemonium processing at the deepest level of human cognition, *see generally* STANISLAS DEHAENE, READING IN THE BRAIN (2009) and in particular, pp. 46–51 and 122–69.

[16] *See* DENNETT, *supra* note 12, at 300–03 (1991).

[17] Weather prediction and other applications that depend on statistical modeling of extremely large amounts of data are one such class of problems. There are of course many many others.

[18] For a reflection on the near-universally accepted but often overlooked necessity of stringently separating value judgment claims from empirical fact claims, *see* D. Michael Risinger, *Unsafe Verdicts: The Need for Reformed Standards for the Trial and Review of Factual Innocence Claims*, 41 HOUSTON L. REV. 1281, 1290–1301 (2004). See also Mark P. Gergen, *The Jury's Role in Deciding Normative Questions in the American Common Law*, 68 Fordham L. Rev. 407 (1999).

7. Attempts at formalized (or symbolized but only semi-formalized[19]) models of human decision such as expected utility theory are deficient in that they solve the normative problem by delegating it to the individual user, who selects and subjectively mathematizes the value judgments about various decisions. While this might provide useful guidance to an individual, mostly by making them consciously aware of and responsible for their own explicit value choices and weights, it is a poor vehicle for use in any social enterprise like law where individual decision makers are neither theoretically nor morally free to select utilities, nor to assign weights to them.[20]

---

[19] Symbolization often occurs when people are working in the hope of developing a fully specified formal system without actually having accomplished that as yet. Symbolization is also helpful to designate variables or parameters in well-established functional relationships independent of full formality, when actual numbers can defensibly be supplied for the parameters or variables in many or all individual cases. But in some cases, such as the Hand formula set out as an epigram above, the symbolized expression seems to be more of a pretension to mathematics than anything for which symbolized expression adds any real value. In some areas, symbolization is more of a professional tic than anything else. This seems especially true in some areas of philosophy after analytic philosophy's early 20th century domination by symbolic logic and what came to be called "logical positivism." Even after the subsidence of logical positivism, the tic appears to remain. The eminent philosopher of science Larry Laudan tells a joke illustrating the phenomenon, beginning "If a chicken C crosses a road R…." (Personal communication.) Sometimes this kind of usage becomes no more than a kind of abbreviation, like an acronym, but it is often harder to remember the symbolized abbreviation than it would be to repeat the word or words involved, so that no efficiency is gained.

However, I think the worst abuses of feckless symbolization occur in the literature generally referred to as "law and economics," and far from being merely an unnecessary irritation, this symbolization can actually cover up the weakly warranted claims about relationships between variables, as I discuss more fully in note 22 *infra*.

[20] I believe this is taken to be an obvious proposition in most legal literature, and especially in the literature dealing with the policy implications of empirical findings. The Supreme Court has said as much. For instance, more than half a century ago, in discussing the concept of discretion in administrative rulemaking, the court said "The 'discretion' protected by the section is not that of the judge--a power to decide within the limits of positive rules of law subject to judicial review. It is the discretion of the executive or the administrator to act according to one's judgment of the best course, a concept of substantial historical ancestry in American law. (Footnote omitted)." Dalehite v. United States, 346 U.S. 15, 33 (1953). And consider the following exposition in an opinion dealing with the discretion to make policy value judgments when making rules, and the lack of such authority when applying such rules (the specific subject is whether the government violated its own regulations in releasing for administration to the public a batch of live virus poliomyelitis vaccine which induced the disease in the plaintiff):

> In this case, the nature of the rules is not attacked, but rather the way the rules were applied. The history of the process of regulation-making in this case contains no implication that there was any intended delegation, should a new context arise, that D.B.S. [Division of Biologic Standards] should consider not only the outcome of applying the regulations honestly, but also the possible desirability of changing the regulations, *sub rosa,* to a standard of lower safety because of a judgment of net gain in the public welfare to be gained therefrom in the minds of the personnel charged with enforcing the regulations. Further, the fact that lots began to fail under the standard imposed by the regulation can hardly be called a "new circumstance." If the concept of safety regulation is to have any meaning, the possibility of failure, even widespread failure, must be taken as contemplated by the regulation. The discretion to revise a safety regulation in the direction of greater public risk should not be assumed absent clear language supporting such a conclusion in the regulation itself. There was none in this case. The public may not be subjected to the now-you-see-it, now-you-don't protection of safety regulations that a finding of discretion in this case would create.
>
> The increased risk imposed upon the public by the disregarding of 73.114(b)(1)(iii) was statistically small. The regulations could properly have been promulgated in such a way as to impose that risk on

8. Any process that results in subjective generation of specific numbers attached to either probabilities of facts in the world (much less of value judgments as "utilities") is deeply suspect, in that the numbers so generated cannot properly be specified as numerical points, but only at best as ranges, and even the ranges may be heavily contestable.[21]

9. Symbolized and semi-formal systems with variables that cannot be mathematized with functional precision may have interest as games, or as hoped-for way stations on the path to something better, but when promoted as fit for practical applications in legal and other social processes, they often become vehicles of mystification. When supplied values are more specific than the underlying data will support, the resulting numbers may become inappropriate, or inappropriately unquestioned, proxies for reality. Vide the Herfindahl-Hirschman index, and much of law and economics.[22]

> the public had the initial judgment been that the benefits coming from such a course justified it. This was not the case, however. The regulations were promulgated by a procedure established to give hearing to any who might object to the balancing done by the agency, and once promulgated have the force of law and reflect the proper balancing between risk and benefit in the judgment of those lawfully empowered to make that judgment according to the process of the law. That judgment may only be changed by the process of the law. Any deviation from the regulations which imposes a higher risk on the public, no matter how slight, is not a matter of discretion. It may not be done.

Griffin v. U.S., 351 F. Supp. 10, 32–33 (E.D. Pa. 1972) (Newcomer, J.). In the spirit of full disclosure, I should note that that I was Judge Newcomer's law clerk when this case was decided.

Thus the administrative authorities had a normative warrant when making regulations but little or no such warrant when applying the regulations. An important issue is thus whether individual decision makers such a judges or juries have been delegated such normative authority through an explicit or implied normative warrant in a decisional standard. This raises the whole Rules vs. Standards debate that has occupied so much of jurisprudential writing in the past two decades. For a useful conceptual roadmap see Gergen, supra note 18, especially the extended footnote 1.

[21] Professor Franklin suggests that in these circumstances the variables in formalized expressions ought at least to be expressed as functions covering ranges of values, or in some other indeterminacy form. *See* Franklin, *supra* note 7, at ____.

[22] I first became sensitized to this problem in the early 1980s when cogitating on the difficult notion of concentration of market power, a central concern of anti-trust law. To make a long story short, the Hirfindahl-Hirschman index was (and is) a mathematized expression supposedly reflecting market concentration. One of the terms in the expression asserts that the strength of the concentration is related to the square of one variable. When I first saw the expression I immediately said to myself (and to those who happened to be present) "why squared?" meaning, what is it about the world that makes the notion of market concentration depend on the relationship represented by the exponent "2" rather than "3" or "1.5" or any other possible exponent? My friends and colleagues Charles Sullivan and Neil Cohen were there at the time, and, being more interested in anti-trust law than I was, took the question and ran with it (giving me a nice hat-tip for having framed the question). They published a fine article in the Texas Law Review showing, inter alia, that the selection of the exponent was essentially a policy judgment, not an empirical fact. See Neil B. Cohen & Charles A. Sullivan, *The Herfindalh-Hirschman Index and the New Antitrust Merger Guidelines: Concentrating on Concentration*, 62 TEX. L. REV. 453 (1983) (see n. 158 for the hat-tip).

The general point here is well summed up by John Ziman: "In many cases, the crudity and unreality of the assumptions are hidden under a mass of mathematical formulae, thus mercifully rendering a dubious argument completely opaque." Ziman, *supra* note 5, at 172. Lest one dismiss Ziman as a math phobe, one should recall that he was a theoretical physicist and a Fellow of the Royal Society.

10. Limiting ourselves to formal models of solely factual inference eliminates some problems but not others. For instance, it eliminates, at least notionally, the problem of value judgment. It does not eliminate the problem of indeterminate values of variables.

11. There are a limited number of functions (in the non-mathematical sense) that can be suggested for any proposed formalized approach to inference.

12. One such proposed function is that a formalized approach to inference can describe inference in such a way that it has operationalizable advantages over humans performing the same function. This is clearly true for some inferential domains, such as predictive meteorology.[23]

13. When one moves away from physical systems fairly directly describable through quantified measurement systems of established natural sciences such as physics and chemistry, the quantitative approach breaks down.

14. The inferential domain furthest from the optimum preconditions for quantitative modeling is reconstruction of the facts of a human episode in the past which is the basis of some controversy that might be presented for legal resolution.

15. To date, no formal attempt at describing such inferential processes, including various fuzzy logic applications and forms of Bayesian probability analysis, is sufficient to propose as a programmable mechanism to replace humans in regard to pure factual inference regarding human episodes, even if that were a function properly separable from value judgment functions in legal systems.

16. Given the difficulties of determining and gaining agreement on the appropriate level of granulation for describing human episodes (the inferential detail taxonomy problem[24]), and

---

[23] *See* IWO BIALYNICKI-BIRULA AND IWONA BIALYNICKI-BIRULA, MODELING REALITY: HOW COMPUTERS MIRROR LIFE 147 (2004) for a list of such applications. However, even with these models there is a problem with directionality that may interfere with inference models in the legal setting. In the legal setting, inference often proceeds from effects to causes. As meteorology shows, this can present special problems, because more than one initial set of causes can end up with indistinguishable effects. Can meteorologic models take local observations today and reconstruct the weather conditions at remote places ten days ago, as accurately as they can take the remote conditions ten days ago and predict the weather locally today? Similar issues pervade some areas, such as medical diagnosis, *see e.g.*, Nancy S. Kim and Frank Keil, *From Symptoms to Causes: Diversity Effects in Diagnostic Reasoning*, 31 MEMORY & COGNITION 155 (2003), and law generally, and must be dealt with in any inferential system with legal applications.

[24] That is to say, at what level of granularity are propositions about the facts of a human episode to be incorporated into the model as nodes in a network of inferential relations. Too gross and important information is submerged or lost, too fine and the tool is too burdensome to be used in real world case settings in real time. In addition, when reasonable people can differ about how to characterize the strength of connections between nodes (which is most of the time), any system with enough nodes to deal with real episodes will often have outputs so indeterminate as to be of no value. How to tame this by deriving connective strengths that command general assent in real cases is an intractable problem. See Dale A. Nance, *The Inferential Arrow: A Comment on Interdisciplinary Conversation*, 10 L. Prob. & R 87 (2007). And this is only in regard to propositions of fact, without the complicating factor of warranted normative judgment discussed in proposition 18 below.

the difficulty of determining any defensible and sufficiently specific numbers for the base rate occurrence (and therefore inferential strengths between facts) for the vast majority of the detail variables (no matter how characterized), a formal system that could be operationalized to yield superior results to humans in regard to such a task is perhaps unachievable, even in principle.

17. Even if proposition 16 is not true in principle, it is practically true for the foreseeable future, and thus operationalized formal modeling could not replace humans any time soon, even excluding the value judgment problem.[25]

18. The value judgment problem is likewise intractable as long as the reality of legal decision making (perhaps unlike some of the less self aware accounts of it) inevitably involves, not mechanical reconstruction of actual "facts in the world" with a mechanical application of decision rules to that specified set of empirically justified propositions, but in many cases involves a normative warrant given by society to the individual decision maker, which is both broad enough to require the consideration of the value of individual outcomes under general criteria of value, and sufficiently narrow not to leave the decision maker entirely free to provide his or her own values. In such a situation, the "factfinders" must handle the details of the empirical facts themselves in a thickly descriptive way to inform the proper fulfillment of the normative warrant.[26] As a result, a mechanized formal system could not replace humans even if it were superior at factual inference simpliciter.

19. If a formal system cannot replace humans, perhaps the formal system can be taught to legal decision-makers sufficiently to guide decisions.[27] But this too seems questionable.

20. Natural language comes more easily to the majority of humans than facility with formal systems. *Vide* the difficulties of giving people competence with even basic mathematics even when taught merely as a system of basic computational operations. When we hit algebra, the penetration of facility plummets. Formal systems are literally a foreign language to a high percentage of people, with only a minority having a high degree of natural aptitude[28] (this also

---

[25] To be fair to the AI people who were at the conference in Pittsburgh which generated this and the other papers, their general response to these last two points was that their work was not directed toward the creation of inferential systems that could replace humans in the legal setting in regard to any legally relevant inferential task.

[26] *See* the discussion in Risinger, *supra* note 15, at 1295–1301.

[27] This is the position taken by most of the AI people at the conference.

[28] This is, of course, a contestable claim. For instance, Keith Devlin argues that the aptitude for mathematical thinking is more widespread than is often supposed. But even he concedes, as he must, that symbolization often creates conditions that interfere with, rather than foster, underlying mathematical aptitudes. Vide the results on various logically equivalent presentations of Watson's test, some in natural language narrative form, some more symbolized. *See* KEITH DEVLIN, THE MATH GENE 115–17. Although Devlin's account of why this is so is perhaps overly optimistic in regard to the general capacity to move from concrete reasoning to symbolized reasoning "about abstract objects that bear virtually no relationship to anything in the real world," *id*. at 268, he also concedes that humans display widely varying aptitude for mastering the special skill of comfort in a world of abstracted symbols, and that for most people, unlike basic natural language competence, this requires very hard work. In this way mathematics is more like music than it is like natural language.

seems difficult for those with a high degree of aptitude to appreciate). If formal systems had to be mastered to be a legal decision maker, it would undermine participation of most of the citizenry on juries and change radically the selection criteria for lawyers and judges.

21. However, formal models of inference can provide insights into the kinds of inferential errors that ordinary reasoners may fall into. This is the same role traditionally ascribed to deductive logic for deductive reasoning, but since any modern account of inferential reasoning must take into account probability in some form, identifying common probabilistic errors, such as ignoring base rate considerations, seems a proper function for formal exploration of inference.[29]

22. However, once identified, such dangers must be communicated to legal decision makers effectively in natural language. This translation and communication problem is one great challenge for the immediate future.[30]

23. Another great challenge in regard to properly integrating the information derived from formal systems into the legal decision making process, comes about when well-warranted sciences such as DNA typing generate properly mathematized characterizations of the results of some process. Guidance on how to combine this with the non-mathematized evidence of the case, given intelligibly in natural language, without generating a specific net prior probability from the non-quantified evidence of the case, is a great challenge not yet satisfactorily worked out.[31]

And so ends my screed against feckless symbolization in legal analysis or in proposals for practice. But there are a couple of themes in it that I will have to expand upon a bit in order to lay the groundwork for what I have to say about the meaning of standards of proof, and the problem of communicating those meanings to jurors in such a way that they decide in accordance with those intendments. It should be obvious by now that I am of the school of

[29] This is the general program of the "heuristics and biases" school of research, that is, to compare human approaches to reasoning under uncertainty to the idealized formal models of rationality prescribed for such reasoning. Whether the departures manifested by actual humans are or are not to be taken as "errors," and under what conditions, is the subject of ongoing and often heated debate between members of the "heuristics and biases" school of thought, and members of the "fast and frugal heuristics" school of thought. *See generally* MARK KELMAN, THE HEURISTICS DEBATE (2011).

[30] I am convinced that judges and juries more easily understand illustrations using ordinal relations and equivalence when considering the main lessons of likelihood ratios, for instance, than they do cardinal expressions. "You would be just as likely to see that piece of information if the defendant was innocent as you would if the defendant is guilty" is the kind of thing that is easily grasped.

[31] See the reflections in Janet Chaseling, *Statistics in Court: The Ultimate Communication Problem*, PROCEEDINGS OF THE SEVENTH INTERNATIONAL CONFERENCE ON TEACHING STATISTICS (ICOTS 7), 2006.

thought who thinks that attempts at formalizing the meaning of burdens through the use of numerical expressions of probability is probably a bad idea in theory and definitely a bad idea in practice. Formal mathematization can have all sorts of unintended consequences. Take, for instance, the various attempts to leverage the so-called "Blackstone ratio" into a formal expression of the numerically expressed probability lying behind the notion of proof beyond a reasonable doubt. As Professor Allen and Professor Laudan have shown to a fare-thee-well, the ratio can be as well realized by raising the number of the guilty defendants acquitted as by lowering the number of innocents convicted.[32] And beyond that, the ratio in its Blackstone form can apparently easily be taken to imply the acceptability of a nearly 10% rate of factually false convictions.[33] Put that way, I suspect it would seem wrong to a majority of people. The trip from a ratio expression to a percentage to a probability expression to a normative expression about system performance is a long journey over a path fraught with pitfalls. [34]

---

[32] Ronald J. Allen and Larry Laudan, *Deadly Dilemmas*, 41 TEXAS TECH L. REV. 65, 75-79(2008).

[33] I say "apparently" because I have often seen this done, even thought Blackstone's ratio on its face deals with a ratio of the convicted innocent to the acquitted guilty, not the convicted innocent to the convicted guilty.

[34] As I have previously written:

> As for the "Blackstone ratio," that is the label now generally given to Blackstone's version of the moral assertion "it is better that ____guilty go free than that one innocent be convicted." I believe that all we can say about the intendment of this expression was that it was meant as a general declaration that, for any given crime, an error that convicts an innocent person is much worse morally than an error that acquits a guilty person. The number Blackstone chose to make this point was ten, although Alexander Volokh has rather amusingly shown that the notion is ancient, and that various thinkers over the centuries have put the number in the blank at various values between one and a thousand. see Alexander Volakh, *n Guilty Men*, 146 U. PA. L. REV. 173(1997).
>
> The trope first appeared in English in Sir John Fortescue's *De Laudibus Legum Angliae*, composed in the late 1460s, where the number employed was 20. Sir Matthew Hale used the formula in the 1670s, but his number was only 5. What is clear is that when Blackstone used the ratio image, like Hale and Fortescue before him, he was not speaking as a mathematician or probability theorist or formal logician, but as what he was—a practical lawyer/judge commenting on the subjective certainty that should be required of juries making findings of guilt in individual criminal trials. It was a way to justify the principle of calling on jurors to apply great caution in convicting individual defendants in particular cases when the specific evidence of their guilt was relatively weak. It seems unlikely that the idea of actually determining the number of true acquittals, false acquittals, true convictions and false convictions produced by the criminal justice system as a whole, and actually comparing their ratios in any combination as a measure of system performance, would have struck him as possible or useful, had he ever formulated the question of system performance in that way, which so far as I know he never did.
>
> So the Blackstone Ratio as analyzed by Allen and Laudan is in a central sense not really the Blackstone Ratio. Nevertheless, for those who have been tempted to convert the Blackstone Ratio to a canonical measure of system performance, Allen and Laudan establish how unconstraining the Blackstone Ratio used as a measure of system performance really is, since, taken as a

Nor can any other approach to mathematization currently supply credible numbers to mathematize what is meant by standards of proof. The main such approach seems to be subjective expected utility decision theory or some other aspect of modern Bayesianism, but in fact I believe such approaches are non-starters for a number of reasons, including the problem of value judgment indicated in the Against Symbolization text, but I should expand upon these reasons a bit more for current purposes. In doing so, I will be drawing liberally from another recent article in Law, Probability & Risk,[35] for the initial four or so paragraphs, but I blessedly won't be reproducing the whole thing.

I lived through the "Bayes Wars" of the 1980s and 90s which were fought over the general theory of evidence and inference.[36] But I was more an observer than a participant, and at least to me those debates seemed more driven by different visions of human capacities than by full-scale commitment to "Bayesianism" (in quotations) as the only revealed way to rational inference from evidence. And there is no doubt that the term "Bayesianism" now covers schools of thought that would have surprised the Reverend Bayes himself,[37] and which embody a set of doctrines at least some of which are not implied by Bayes's Theorem itself.

---

mathematical expression, its conditions can formally be met by holding the number of convicted innocents steady and raising the number of acquittals of the guilty.

D. Michael Risinger, *Tragic Consequences of Deadly Dilemmas: A Response to Allen and Laudan*, 40 Seton Hall L. Rev. 991, 1002-1003 (2010).

[35] D. Michael Risinger, *Reservations About Likelihood Ratios (and Some Other Aspects of Forensic" Bayesianism")*, awaiting print publication but available online at oxfordjournals.org.

[36] For a prècis, see D. Michael Risinger, *Introduction to "Bayes Wars Redivivus,"* 8 INTERNATIONAL COMMENTARY ON EVIDENCE, article 1.

[37] The exposition in the single article on the subject attributed to Thomas Bayes, published posthumously in 1763 (after editing by Richard Price, and with an introduction and commentary by him) in the Philosophical Transactions of the Royal Society, set out certain important aspects of what is now known as Bayes's Theorem, without any specific algebraic formulae. *See* (as the title quaintly puts it) *An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S.*, 53 PHILOSOPHICAL TRANSACTIONS 370 (1763), available at http://rstl.royalsocietypublishing.org/content/53/370 But it was clearly Laplace, almost certainly independently, whose work gave the more generalized notion its impetus and content. *See generally* STEPHEN M. STIGLER, THE HISTORY OF STATISTICS: MEASUREMENT OF UNCERTAINTY BEFORE 1900 99–138 (1986). As Stephen Fienberg has noted (noting others who have noted the same thing), "Bayes did not actually give us a statement of Bayes' Theorem, either in its discrete form (this came with Laplace), or in its continuous form with integration, although he solved a special case of the latter." Stephen E. Fienberg, *When did Bayesian Inference Become "Bayesian"?* 1 BAYESIAN ANALYSIS 1, 3 (2006) (formula and note omitted).

But let me not put forth such a perhaps controversial and contentious proposition on my own authority. Let me quote the eminent statistician, the late Jack Good, an early adherent, who said only a few years ago that the term Bayesian is now usually used to refer to "a whole philosophy or methodology in which 'subjective' or logical probabilities are used."[38] So Bayes's Theorem does not define modern Bayesianism (or "neo-Bayesianism" as it is sometimes called) so much as the mysterious concept of "subjective" probabilities[39] does (that is, mysterious at least in some ways to me[40]). And indeed, a little research reveals what I am sure many readers already know, that there are "objective" applications of Bayes's Theorem, so-called "frequentist" applications of Bayes's Theorem, and so forth,[41] and these would hardly qualify for a seat at the

[38] Quoted from Good's personal correspondence with Stephen Fienberg in Fienberg, supra note 7, at 15.

[39] Professor Fienberg traces the genesis of the explicit notion of "subjective" probabilities as degrees of belief to John Maynard Keynes's 1921 TREATISE ON PROBABILITY, although he notes that Keynes "allowed for the possibility that degrees of belief might not be numerically measurable." Fienberg, *supra* note 6, at 9. Fienberg also notes the use of the notion of "degrees of reasonable belief" (the qualifier seems later to have fallen out) by Harold Jeffreys and Dorothy Wrinch in a 1919 paper. Later, beginning with Borel and Ramsay, approaches to quantifying subjective probability through mind experiments involving betting under various assumptions and constraints were put forth as adequate mechanisms for quantification.

[40] I am aware of one sense in which all probabilities must be regarded as, in that sense, subjective, in the same way that language, being an abstraction, is in the same sense subjective. Both are products of mind, and are therefore mind-dependent. Something like this seems to have been the starting point for Bruno De Finetti. See BRUNO DE FINETTI, THEORY OF PROBABILITY, v.1, p.1 (English translation, 1974): "Probability does not exist," (quoted in Robert F. Nau, *De Finetti Was Right: Probability Does Not Exist*, 51 Theory and Decision 89 (2001)). Perhaps the most mysterious thing, then, is why anyone claims any probabilities are <u>not</u> subjective. However, I think I get the main point of the frequentists: despite various levels of mystery (including the ontological status of mathematics in general, see Mary Leng, MATHEMATICS AND REALITY, Oxford U. Press, 2010), objective or "frequentist" probability is the appropriate model when one has reason to believe that abstract notions of exact mathematical probability specification (represented in their simplest form by specified "mental experiment" urn drawing problems or similarly specified mental experiments regarding gambling games) actually represent a reasonable model of the empirical world in which a problem is faced and must be answered (real poker games, etc.). When such circumstances do not exist (and more often than not they don't) best guesses (subjective probabilities) must be resorted to, or else we must defer decision.

But in both language and in notions of probability, there is a necessary contribution from an exterior physical reality, and more so in regard to warranted probability statements about factual conditions than in regard to some kinds of natural language expressions such as value statements. So my own emphasis is not on the subjective nature of probabilities *vel non* in this general sense, but on the empirical warrant that is, or ought to be, necessary to derive and use a probability statement in different contexts. It may be that numbers derived from subjective individual mental betting experiments are perfectly sufficient for some kinds of problems (contexts in which no iterative process is available to refine prior assumptions out of results, a decision must be made, and all the costs will fall on the betting party or parties, for instance), but not in others. However, appropriate criteria for acceptable best guesses for different classes of problems seem hardly ever to be discussed, at least so far as I have been able to discover. That is the most mysterious part about subjective probabilities, to me.

[41] In a recent popular book on Bayesianism, Sharon Bertsch McGrain attributes to Jack Good the tongue-in-cheek claim to have identified 46,656 varieties of Bayesians. She then lists the following varieties: subjective, personalist, objective, empirical, semi-empirical, semi-Bayes, epistemic, intuitionist, logical, fuzzy, hierarchical,

table among today's Bayesians, many of whom appear to be committed to the celebration of subjective probability without much attention to the issue of varying the belief warrant required for the assignment of those probabilities in different classes of decision problems.[42]

There seems to be a consensus that neo-Bayesianism, or simply "Bayesianism" full stop, as it is now known, as a "whole philosophy or methodology" began to emerge as a movement in the 1950s, and came to full flood in the 60s with, among other things, the emergence of influential rational choice models of expected utility decision theory.[43] These models, perhaps best represented by the work of Howard Raiffa and Robert Schlaifer,[44] have plenty of helpful applications, but they may not be as universalizable as their proponents believe.[45] But nevertheless, Bayesianism based on freewheeling notions of subjective probability became a constellation of beliefs and investments that attracted many adherents. In this, it reminds me somewhat of another movement of a hundred years earlier which I have given some study to (and written about a bit), that is, August Comte's Positivism, which was driven to a similar secular desire to identify and embody pure rationalism, and which ironically turned into a kind of secular religion.[46]

---

pseudo, quasi, compound, parametric, nonparametric, hyperparametric, and non-hyperparametric Bayes. See SHARON BERTSCH MCGRAIN, THE THEORY THAT WOULD NOT DIE: HOW BAYES' RULE CRACKED THE ENIGMA CODE, HUNTED DOWN RUSSIAN SUBMARINES & EMERGED TRIUMPHANT FROM TWO CENTURIES OF CONTROVERSY 129 (2011) (e. e. cummings-like lack of initial capitalization copied directly from spine and title page of book). These categories obviously are not mutually exclusive. A further word about this book might be in order. It is an exercise in a kind of triumphalist Whig history, which fails to make very clear the actual contributions of "Bayes" to many of the episodes it recounts as examples of Bayesian triumph. On the other hand, it is a terrific compendium of anecdotes concerning the contempt in which "frequentists" held "Bayesians" (and vice versa) in the statistics community for over half a century. I have blessedly never seen such vitriol in the circles I inhabit. Who would have guessed that statisticians could be so much more vicious than law professors in any context.

[42] See discussion in fn. 10 *supra*. In fact, in many scientific contexts, "subjective" priors are not assigned by a subjective process at all, but according to formulas that appear to give reasonable prior assumptions for classes of cases. *See* Robert E. Kass and Larry Wasserman, *The Selection of Prior Distributions by Formal Rules,* 91. J. AM. STAT. ASSOC. 1343 (1996).

[43] Fienberg, supra n. 7 at 14-20; McGrain, supra n. 11 at 97-107.

[44] The wellspring work is HOWARD RAIFFA & ROBERT SCHLAIFER, APPLIED STATISTICAL DECISION THEORY (1961).

[45] *See generally* Matt Jones and Bradley C. Love, *Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of the Bayesian Model of Cognition*, 34 BEHAVIORAL AND BRAIN SCIENCES 169 (2011).

[46] *See* D. Michael Risinger, *Boxes in Boxes: Julian Barnes, Conan Doyle, Sherlock Holmes and the Edalji Case*, 4 INT'L COMMENT. ON EVIDENCE, Iss. 2, Art. 3 at 7, n. 13, *available at* http://www.bepress.com/ice/vol4/-iss2/art3.:

While the formal expression of Bayes's Theorem (which was generated by LaPlace) is no doubt an analytic truth which can be a powerful model of rational inference whenever well warranted quantification is available in regard to information being added to a system (like serial proffers of evidence directed to a single issue at trial), it is of little use without such numbers, since both the initial characterization of the starting point (the "prior probability") and the updating by new information, only work in any helpful way with numbers. But when there is no fully specified formal system (like the artificial worlds of gambling games or urn drawing experiments) to provide numbers, and when there is no real empirical data to provide numbers, what to do? The Bayesian inference machine won't work without numbers. The modern Bayesian's answer is, we will just make them up.

Well, not exactly, but almost. In the 1920s Frank Ramsay of Cambridge, perhaps influenced by Emile Borel, proposed a way of generating numbers that he claimed would then work for probability and Bayesian inference purposes every bit as well as the frequency numbers that were a normal product of the analysis of formally specified systems like games. He said that a person could conduct a mental experiment involving betting on a proposition of fact, and that the honest answer to what you would wager represented a number that then could be used in Bayes's Theorem properly.

Once this process is reflected upon, there are a variety of available objections. First, the resulting numbers are not really the specific probability statements that the numerical values resulting from the forced choice of the mental experiment suggest (they have illusory precision). Second, they are grossly unreliable between persons in many applications, that is, different

---

The foundational theorist of 19th-century positivism was Auguste Comte (1798–1857), who was, among many other things, dedicated to the proposition that the methods of science could be extended to give sure predictive knowledge to human social actions. *See generally* THE POSITIVE PHILOSOPHY OF AUGUSTE COMTE FREELY TRANSLATED AND CONDENSED BY HARRIET MARTINEAU (1855); *see also* JOHN H. ZAMMITO, A NICE DERANGEMENT OF EPISTEMES: POST-POSITIVISM FROM QUINE TO LATOUR 6–8 (2004). Comte, like Marx, was an enemy of religion who rather paradoxically founded a sort of social religion. While Comte's positions are looked upon as foundational to the social sciences, his views influenced popular views of science more than they influenced science itself. Ironically, he wrote at a time when the switch among natural scientists from certain knowledge goals to probabilistic goals was just beginning to take hold. *See* LARRY LAUDAN, SCIENCE AND VALUES 83–85 (1984).

people will give different numbers. Third, when used for deriving probability statements about facts in the world, the relationship of even the average of multiple bets by multiple people may have a weak bearing on the underlying empirical reality. And even this operation may in some sense be out of bounds, because many Bayesians will not necessarily go past saying that an individual's actions based on their beliefs (as measured by their bets) should obey the probability calculus, including Bayses's Theorem, even if those beliefs are wrong or ill-founded. All this may be fine for personal uses in defining personal utilities in decision circumstances where the costs of being wrong will fall on the person making the subjective bets, but not otherwise. Additionally, using a betting model assumes people are familiar with betting and skilled at making real wagers. This is hardly a social given. And there are other available objections.[47]

So how what does this have to do with standards of proof? One reason some Bayesians think standards of proof are a good target for Bayesian mathematization is that standards of proof, whatever else they are for, are intended to speak to jurors about their level of certainty concerning the material issues of the case they are deciding after they have seen the evidence produced at the trial. Like any such individual personal level of certainty about specific factual issues (and I will limit myself for the time being to actual binary "facts in the world," the factual details of the episode that gave rise to the litigated controversy), these are subjective states which do not in themselves directly reflect a mathematized, or perhaps even mathematizable probability directly (unlike intelligent evaluations of the state of play in a well-specified and well understood gambling game). Rather they are best described by the term "degree of belief," a usage first put forth, apparently, by Jeremy Bentham,[48] but which re-entered philosophical and mathematical consideration of such things in the late teens and early 20s of the Twentieth Century, as I described in fn. 37 above. In fact "…'degree of belief is the central notion of subjective

[47] For a catalogue of objections to the betting model for number generation, see Lina Eriksson & Alan Hajek, *What Are Degrees of Belief?* 86 STUDIA LOGICA 183, 186–90. One should note that Ramsay and later followers regarded the bet as a proper number only if one were willing to act on it, but what that would mean in many contexts where subjective probabilities of this sort are used is difficult to say.

[48] "Nobody can be ignorant, that belief is susceptible of different degrees of strength, or intensity." Jeremy Bentham in JEREMY BENTHAM, A TREATISE ON JUDICIAL EVIDENCE (1825), p. 40. This is from the one volume work edited by M. Dumont containing Bentham's views on evidence, not the five volume work on the same subject edited by Bentham's protégé John Stuart Mill and published two years later. The statement given holds true as much for normative or metaphysical statements as for statements concerning things properly characterized as factual, but we will come to that in due course.

probability theory, the core concept around which the entire theory is built."[49] What the betting model does for modern Bayesians is provide numbers which they claim measure degrees of belief, that can then be fully mathematically operationalized.

I am skeptical of this approach for a couple of reasons. First, while 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,[50] a purpose for which formalized modern probability notions are ill suited functionally and expressively. [51] This does not mean that we are wrong in common parlance to say that standards of proof involve judgments of how likely, or how probable, the existence of a fact is, but merely that those terms (like Blackstone's use of his ratio) are not properly seen as statements entailing or implying any formal version of modern mathematized probability theory. The notions of "probable" and "likely" predate Pascal by millennia.[52] But my skepticism is driven by something deeper. I

[49] Eriksson & Hajek, supra note 45 at 184.

[50] Jim Whitman has very effectively shown this in JAMES Q.WHITMAN, ORIGINS OF REASONABLE DOUBT: THE THEOLOGICAL ROOTS OF THE CRIMINAL TRIAL (2008), and none of the other historical analyses bearing on the development of legal standards of proof cited in note 50 infra contradict that conclusion.

[51] Until fairly recently, I also thought of standards of proof primarily as devices for distributing error. However, I now think that although such error distribution is an inevitable result of the operation of such standards, the purpose, as opposed to the consequence, of such standards, is to get the jurors to understand the moral burden they bear in the individual case, not to reflect upon or speculate about how this ties in to system error and its distribution. Kevin Clermont's ruminations on standards of proof over a long period of time seem generally consistent with this position, and Clermont establishes it to be the view of much of the rest of the world. See Kevin M. Clermont, *Procedure's Magical Number three: Psychological Bases for Standards of Decision*, 72 Cornell L. Rev. 1115 (1987); Kevin M. Clermont & Emily Sherwin, *A Comparative View of Standards of Proof*, 50 Am. J. Comp. L. 243 (2002); Kevin M. Clermont, *Standards of Proof in Japan and the United States*, 37 Cornell Int'l L. J. 263 (2004); and Kevin M. Clermont, , 33 Vt. L. Rev. 469 (2009). Clermont labels the individual degree of belief required in such setting "confidence" as opposed to probability. This is a potentially a bit confusing, since "confidence" is also a standard concept in formal probability theory. For an attempt to leverage that formal concept into an explanation for why it is wrong to conceive of the preponderance standard as a knife-edge point estimate of probability which is the least bit greater than .5, see Neil B. *Cohen, Confidence in Probability, Burdens of Persuasion in a World of Imperfect Knowledge*, 60 NYU L. Rev. 385 (1985).

[52] There is a rich literature on pre-Pascal probability notions, the most complete of which is JAMES FRANKLIN, THE SCIENCE OF CONJECTURE: EVIDENCE AND PROBABILITY BEFORE PASCAL (2001); see also WHITMAN, supra note 48; BARBARA J. SHAPIRO, BEYOND REASONABLE DOUBT AND PROBABLE CAUSE: HISTORICAL PERSPECTIVES ON THE ANGLO-AMERICAN LAW OF EVIDENCE (1991), IAN HACKING, THE EMERGENCE OF PROBABILITY: A PHILOSOPHICAL STUDY OF THE EARLY IDEAS ABOUT PROBABILITY (1975) and LORRAINE DASTON, CLASSICAL PROBABILITY IN THE ENLIGHTENMENT (1971). All seem to agree that legal proceedings and theological controversies generated much of

think betting exercises are poor proxy measures of the thing that, to my mind, truly reflects degrees of belief—the intensity of the surprise that would be experienced if it were established that the proposition believed were false.

My central claim is that people believe something to be true to the extent they would be surprised to find out it was false.[53] Another thesis is that this is the best measure of subjective certainty, of the degree of belief, and that a person's own degree of belief is best revealed to his or her self by a different mental experiment—not 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. Finally, I believe that standards of proof are addressed to making jurors determine this degree of belief, and then norming them to understand the degree of belief the law requires for an affirmative finding.

But how do we measure these levels of surprise, or at any rate induce jurors to measure them in themselves. Not by artificially generated numbers to put into a full probability calculus, but by a well ordered system of categories designated by words like what are now called (rather unfortunately, I now believe) "words of estimative probability."[54] In fact, the traditional three standards of proof of our litigation system are a rank-ordered, three-category system defined by various formulas which can be thought of as words (or word formulas) of estimative probability. But what I propose is a system which uses what I would prefer to call "words of estimative surprise," such as mildly surprised, surprised, greatly surprised, shocked, etc.

---

the non-mathematized thought on certainty and uncertainty. Nor was it all informal, as the Romano-canonical system of proof through categories amounting to proof fractions, with the requirement that the evidence mount up to "full proof" illustrates. This even used numbers, in a way, though not in a way we would easily call "mathematization," or would find very congenial today.

[53] The second epigram set out for this piece ("that's funny") was to show surprise accompanying disturbed belief. Apparently my willingness to define belief in terms of such a subjective emotion as surprise makes me a "liberal dispositionist" in regard to philosophical accounts of belief. See entry "Belief" in the Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/belief/ I am not surprised that some others have embraced the centrality of surprise, but I have not encountered it in any of the legal literature on standards of proof or the jury function.

[54] See e.g., Sherman Kent, *Words of Estimative Probability*, originally published (though classified) in Studies in Intelligence, Fall, 1964, now available at https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html ; Frederick Mosteller and Cleo Yountz, *Quantifying Probabilistic Expressions*, 6 Stat. Sci. 2 (1990).

Some interesting things about such scales is that they can be easily understood as rank ordered, and therefore having the transitive property of numbers, but as not having any of the other properties of mathematization. On the other hand, they are inherently imprecise, not merely because they are not as fine grained as numbers, but because they do not represent equal quantities—they are subjective, not merely in the sense that they are mind-dependent, but in the sense that there will be variability, very likely significant variability, between individuals in the strengths of surprise their designations represent. What we need to attain reliability is a reliable way of norming those we want to use the scale into reasonably consistent designations among humans using the scale.[55] Here is where Gladstone and Victoria come to the rescue.

The "story model" of human information processing suggests that story form facilitates human understanding of the meaning of information in many contexts. [56] What Chesterton has done in the little anecdote reflected in the initial epigram that starts this piece is tell a small story that illustrates how one can be virtually certain of certain facts in the world without recourse to mathematical expression. Indeed, the level of certainty reflected in the Gladstone and Victoria story goes well beyond what most people would require as proof beyond a reasonable doubt. My notion is that various degrees of belief might best be illustrated to juries by such illustrative mini-stories which would capture the law's notion of the required level of surprise one should feel were things not as one must find them to be in order to give judgment.[57]

---

[55] This is, I suppose, an attempt to supply what Wigmore said was wanting: "No one has yet invented or discovered a mode of mmeasurement for the intensity of human belief" 9 John H. Wigmore, Treatise on the Law of Evidence in Trials at Common Law '2497(3rd Ed., 1940).

[56] The "story model" is a model of human cognition and inference that suggests that most ordinary humans process the inferential meaning of ordinary circumstantial information best when it is cast into story form, and that the generation of such plausible stories is central to both accurate human inference and persuasion. It may be seen as an extension of schema theory in psychology, but in the legal setting it is based on the work of Nancy Pennington and Reid Hastie, see Nancy Pennington and Reid Hastie, Evidence Evaluation in Complex Decision Making, 51 J. Personal & Soc. Psych. 242 (1986); Nancy Pennington & Reid Hastie, "The Story Model of Juror Decision Making", ch. 8 in REID HASTIE, ED, INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING (1993). Professor Allen was an early adopter, see e.g. Ronald J. Allen, *The Nature of Juridical Proof,* 13 Cardozo L. Rev. 373, 410 (1991).

[57] And it probably goes without saying that I reject any attempt to give expressions of mathematical modeling to juries in order to define standards of proof. For recent arguments (more or less) in favor of such an instruction, see Peter Tillers and Jonathan Gottfried, *United States v. Copeland,* 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond a Reasonable Doubt is Unquantifiable, 5 L., PROB. & RISK 135 (2006); Jack B. Weinstein and Ian Dewsbury, Comment on the Meaning of "Proof Beyond a Reasonable Doubt," 5 L. PROB, & RISK 167 (2006); and Jon O. Newman, Quantifying the Standard of Proof Beyond a Reasonable Doubt:

Here I must make a confession. I do not have a system of such preferred and acceptable illustrative anecdotes to offer. Gladstone and Victoria won't work very well in the modern world, because the social knowledge that makes it so powerful is simply not part of the average juror's social knowledge today. Such anecdotes would have to be carefully crafted to be consistent with such conditions, and given the splintering of such knowledge into various sub-cultures, it seems a daunting task. In addition, I do not think that the only means to norm the levels of surprise representing required degrees of belief is the story form anecdote or parable. Other images, even ones using numbers in a metaphorical and informal way, such as the expression "one in a million," might be serviceable. Even simply the use of the terms of estimative surprise, like "shocked" for a description of reasonable doubt, might do.[58] As a beginning, I am attaching a schematic which I find helpful in organizing my current thinking, if nothing else. It is perhaps best viewed as a discussion starter.

Finally, the title of this piece asked in part "what should we say to them to get what we want?" Up to now I have proceeded as if this "saying" is all to be done by the court through instructions. However, lawyers also address these issues in closing, and it may be that harnessing the story model and the notion of surprise to the end of alerting and norming the jury to the appropriate degree of belief required by the law's standards of proof might best be done in closing argument. At least it might be found a useful approach by those who do not have the burden of persuasion as a way to effectively arguing these points. Further, on the current state of his knowledge, information and belief, the deponent sayeth not.

A Comment on Three Comments, 5 L. PROB, & RISK 267 (2006). For a critique with which I largely agree, see James Franklin, *Case Comment—United States v. Copeland, 369 F. Supp. 2d 175 (E.D. N.Y. 2005): Quantification of the "Proof Beyond Reasonable Doubt,"* 5 L. PROB, & RISK 267 (2006).

[58] And of course, none of this deals with the last-step value judgment in regard to elements upon which the jury has a normative warrant. The best I can do in regard to such issues is to say that the required degree of belief must apply to the factual predicates upon which the final normative evaluation is rendered.