1991

Credal Probability

David H. Kaye

Penn State Law

Recommended Citation

David H. Kaye, *Credal Probability*, 13 *Cardozo L. Rev.* 647 (1991).

CREDAL PROBABILITY

*D.H. Kaye**

A COMMENT ON PAUL BERGMAN AND AL MOORE'S "MISTRIAL BY LIKELIHOOD RATIO: BAYESIAN ANALYSIS MEETS THE F-WORD"

"Rabbi" Bergman and "Reverend" Moore ("Bergman & Moore") doubt that "ideal triers of facts" would "be Bayesians."[1] Invoking the majesty and mystery of the due process clause, they posit that ideal jurors "must make individualistic judgments about how they think a particular event (or series of events) occurred."[2] Thinking that "Bayesian methodology [compels] triers to make frequency assessments, not believability judgments,"[3] while the law requires them "to make believability judgments at trials,"[4] they conclude that Bayes' rule (and hence the whole apparatus of probability theory), while "relevant to certain trial tasks,"[5] nevertheless "fails as a theoretical factfinding model."[6]

Generations of statisticians, philosophers, and logicians have discussed probability as a measure of belief in empirical propositions and of the validity of inductive arguments.[7] These individuals, I daresay, would find the thesis that probabilities cannot measure or grade beliefs puzzling if not shocking. An extensive literature on inductive

* Regents Professor, Arizona State University College of Law.

1 Bergman & Moore, *Mistrial by Likelihood Ratio: Bayesian Analysis Meets the F-Word*, 13 CARDOZO L. REV. 589, 591 (1991).

2 *Id.*

3 *Id.* at 596; *see also id.* at 595 n. 29 ("Bayesian methodology . . . produces a subjective frequency assessment"); *id.* at 596 n. 32 ("Bayes' Theorem produces frequentist assessments").

4 *Id.* at 594-95.

5 *Id.* at 592 n. 16.

6 *Id.* at 592. Inasmuch as Bergman & Moore concede that likelihood ratios are "well suited" to modeling the rules of evidence involving relevance and probative value, *id.* at 614, it is difficult to know what practical consequences follow from their thesis. They seem to object to a juror's relying on Bayes' rule to come to a belief, but not to relying on legislative or judicial use of conditional probabilities in formulating or applying presumptions and admissibility rules.

7 *See, e.g.*, C. HOWSON & P. URBACH, SCIENTIFIC REASONING: THE BAYESIAN APPROACH (1989); B. SKYRMS, CHOICE AND CHANCE: AN INTRODUCTION TO INDUCTIVE LOGIC (3d ed. 1986); A.W. BURKS, CHANCE, CAUSE, REASON: AN INQUIRY INTO THE NATURE OF SCIENTIFIC EVIDENCE (1977). By an inductive argument, I mean an argument with premises that support a conclusion to some degree, even though these premises can be true while the conclusion can be false. In contrast, a deductive argument must have a true conclusion if its premises are true. *See* B. SKYRMS, *supra*, at 6-13. Factfinding in law (as in all other endeavors) involves assessing inductive arguments and their conclusions.

logic concerns probabilities—variously denominated "credal," "epistemic," "personal," or "subjective"[8]—that, in the judgment of many writers, constitute an acceptable measure of belief in propositions.[9]

Nevertheless, the fact that Bergman & Moore's thesis seems to fly in the face of this received wisdom does not prove it wrong. Furthermore, it is fun to deal with the examples they develop. Part I of this Comment demonstrates how the examples that are supposed to show the "frequentist" character of "Bayesian methodology" or the fallacies in "Bayesian analysis" are easily handled without a frequentist interpretation of probability.[10] Part II very briefly describes one or two reasons to think that an ideal juror's partial beliefs will conform to the calculus of probabilities.

I. HYPOTHETICAL CASES

A. *Sherry's Coin*

Bergman & Moore's first major example of "a subjective frequentist notion of probability"[11] is a dispute between a nine year-old child, Sherry, who tells a parent that a fair coin she flipped three times came up heads each time, and her seven year-old brother, Orin, who insists that it came up heads only twice. Apparently, the parent must decide how many heads appeared. The proper Bayesian analysis is straightforward and revealing. It indicates that the parent cannot believe beyond all doubt that Sherry (or Orin) is correct, and it reveals the partial belief that the parent should have in either story.

Let X stand for the number of heads.[12] Let R be the event that Sherry reported three heads and that Orin reported two.[13] The parent wonders whether $X = 3$ given R. Bayes' rule tells the parent that the

[8] I. LEVI, THE ENTERPRISE OF KNOWLEDGE: AN ESSAY ON KNOWLEDGE, CREDAL PROBABILITY, AND CHANCE § 4.5 (1980); J. POLLOCK, NOMIC PROBABILITY AND THE FOUNDATIONS OF INDUCTION 4 (1990); B. SKYRMS, CAUSAL NECESSITY: A PRAGMATIC INVESTIGATION OF THE NECESSITY OF LAWS IA4 (1980); A.W BURKS, *supra* note 7; C. HOWSON & P. URBACH, *supra* note 7.

[9] *Compare* J.S. MILL, A SYSTEM OF LOGIC: RATIOCINATIVE AND INDUCTIVE ch. 18, § 1, at 535 (8th ed. 1974) ("the probability of an event . . . to us means the degree of expectation of its occurrence, which we are warranted in entertaining by our present evidence") *with* Fishburn, *The Axioms of Subjective Probability*, 1 STATISTICAL SCI. 335, 335 (1986) ("Personalistic views hold that probability measures the confidence that a particular individual has in the truth of a particular proposition").

[10] One could think of a personal probability as an estimate of a hypothetical relative frequency, but this interpretation is strained and unnecessary. *See* Ellman & Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. REV. 1131, 1157 (1979).

[11] Bergman & Moore, *supra* note 1, at 594.

[12] For simplicity, I take it as certain that Sherry flipped the coin exactly three times.

[13] I take this joint event to be certain.

probability that X takes on any particular value x, given the reports that comprise event R, is directly proportional to the product of two probabilities: the "prior probability" that a fair coin would come up heads in x out of three tosses, and the conditional probability, or "likelihood,"[14] of the reports comprising R for the corresponding value x. In symbols:

$$p(x|R) \propto p(x)\,p(R|x) \qquad \text{(A.1)}$$

The prior probability function $p(x)$ is known from the rules governing fair coins. As Bergman & Moore observe, the probability that a fair coin will come up heads three out of three times is $p(3) = 1/8$; the probability of two out of three heads is $p(2) = 3/8$; the probability for one head is $p(1) = 3/8$; and that for no heads is $p(0) = 1/8$.[15]

The likelihood function $p(R|x)$ is more subtle. Since R is a unique event, there is no frequentist probability that can give the parent this number. It follows from the information that Bergman & Moore characterize as "subjective frequentist"—the track records of children like Sherry and Orin—*and* from all the individual circumstances that Bergman & Moore would have the parent employ in making "believability judgments."[16] Based on all such knowledge pertinent to the children's behavior, the parent, let us imagine, lists the probabilities that the children would say what they did for each possible number of heads. The list might look like this:

$$p(R|0) = .01;\ p(R|1) = .02;\ p(R|2) = .20;\ p(R|3) = .77.$$

Substituting these values $p(R|x)$ and those for $p(x)$ observed above into equation A.1 tells the parent the "posterior probability" $p(x|R)$ of each possible number of heads. Notice that $p(3|R) \propto (1/8)(.77) \propto 77$, $p(2|R) \propto (3/8)(.20) \propto 60$, $p(1|R) \propto (3/8)(.20) \propto 6$, and $p(0|R) \propto (1/8)(.01) \propto 1$. Since the sum of the posterior probabilities must be one, we normalize these numbers to find that $p(3|R) = 77/(77 + 60 + 6 + 1) = .53$.

On the available information, the parent cannot be certain of Sherry's story, but can entertain just this degree of actual belief—this partial belief—in her claim that $X = 3$. Whether it is right or wrong,[17] this partial belief in "what really happened"[18] is not

[14] Likelihoods differ from conditional probabilities, but the distinction is not important here. *See* Kaye, *Quantifying Probative Value*, 66 B.U.L. REV. 761 (1986).

[15] Bergman & Moore, *supra* note 1, at 594.

[16] *Id.* at 595.

[17] If the consequences of mistakenly concluding that Sherry's statement is false are neither better nor worse than the consequences of wrongly concluding that Sherry spoke truly, the parent should conclude that the coin always came up heads as Sherry said.

[18] Bergman & Moore, *supra* note 1, at 593.

frequentist. The belief is the posterior probability of three heads given the children's reports as the parent evaluates them in the light of what the parent knows about Sherry, Orin, child development, coins, and anything else pertinent to the dispute.

B. *Taxicabs*

An even simpler analysis disposes of the hypothetical taxicab case that so worries Bergman & Moore.[19] We are told that a cab was involved in a hit-and-run accident in a city in which 85% of all cabs are green and 15% blue, that the only witness (apparently unbiased) testified that he saw a blue cab in the accident, and that tests of this witness under identical conditions show that he identifies blue cabs as blue 80% of the time and green cabs as green 80% of the time.

According to Bayes' rule, the conditional probability that the cab was blue is proportional to the prior probability times the likelihood. That is,

$$p(B|WB) \propto p(B)\, p(WB|B), \tag{B.1}$$

where B stands for a blue cab, and WB for the witness's testimony of a blue cab. Likewise, the posterior probability that the cab was green, given the witness's report that it was blue, is

$$p(G|WB) \propto p(G)\, p(WB|G), \tag{B.2}$$

where G represents a green cab. Dividing equation B.1 by equation B.2 gives the "likelihood ratio" version of Bayes' theorem that Bergman & Moore discuss:

$$\frac{p(B|WB)}{p(G|WB)} = \frac{p(B)}{p(G)}\,\frac{p(WB|B)}{p(WB|G)} \tag{B.3}$$

Assuming that the chance of a cab's being at the accident location and that the degree of care with which a cab is driven are independent of color, the prior odds of a blue cab having been involved in the accident are

$$p(B)/p(G) = 15/85 = 3/17. \tag{B.4}$$

Since the experiment establishes that $p(WB|B) = .8$ and $p(WB|G) = 1 - p(WG|G) = 1 - .8 = .2$, the ratio of the likelihoods is

$$p(WB|B)/p(WB|G) = .8/.2 = 4. \tag{B.5}$$

Substituting equation B.4 and equation B.5 into equation B.3 gives the posterior odds for a blue cab:

$$p(B|WB)/p(G|WB) = 4(3/17) = 12/17. \tag{B.6}$$

The corresponding posterior probability is $p(B|WB) = 12/(12+17)$

---

19 *Id*. at 599.

= 12/29 = .41. The partial beliefs of the ideal juror who relies on this specified information point to a green cab as the culprit.

Bergman & Moore object that this analysis produces a "bizarre result."[20] They imagine a plaintiff who introduces the statistics on the prevalence of blue and green cabs as well as the witness who testifies that the cab he saw was blue—and then argues that since $p(G|WB) = 1 - p(B|WB) = .59$, the culpable cab is green. To reveal what actually causes the difficulty, imagine instead (and perhaps more realistically) that plaintiff merely produces the prevalence statistics (without the witness) to show that the cab is green and argues that since $p(G) = .85$, the case against a green cab should go to the jury and that the ideal juror who follows Bayes' rule must find a green cab liable. Of course, this is just the much mooted problem of naked statistical evidence. The obvious gap in plaintiff's case, which could well justify the opposite verdict, is only exacerbated when plaintiff produces a witness who supplies no link to a green cab and weakens rather than strengthens the naked statistical case.

All this can be, and has been, incorporated into an explicitly Bayesian framework.[21] The juror who understands how Bayes' Theorem works with missing evidence may not be persuaded that simply because most cabs are green, the cab in question was green. This Bayesian juror will be even less impressed with the argument that because most cabs are green *and* a witness reports a blue cab, the cab is green. In short, this ideal juror will avoid any "bizarre result" that follows from a naive or superficial application of Bayesian decision theory.

Still, Bergman & Moore are unhappy with the solution for another reason. They think it assumes that the witness knows that all the cabs are either blue or green and will always make a positive identification.[22] But this too is not a telling objection to the analysis. By hypothesis, there are only blue and green cabs in the city, and the likelihood ratio for the witness's accuracy came from an experiment in which the witness was presented blue and green cabs and responded "correctly" 80 % of the time.[23] If the only "correct" response is a report of "blue" when the cab is blue and "green" when the cab is green, then the likelihood ratio is 4, as shown in equation B.5. If "correct" identifications also include reports of "I don't

20 *Id.* at 600 n. 46.

21 *E.g.*, Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54 (1987); Kaye, *Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?*, 66 B.U.L. REV. 657 (1986).

22 Bergman & Moore, *supra* note 1, at 617.

23 *Id.* at 599 n. 43.

know," then the likelihood ratio involving a report of "blue" has been wrongly computed, but the use of a properly computed likelihood ratio is not impeached. Properly stated, the likelihood ratio must involve the conditional probability that the witness (who knows whatever he knows about the color of cabs in the city) reports "blue" given a blue cab relative to the conditional probability that the witness reports anything else (including an "I don't know" or an "I can't remember") given a blue cab.

Now, if there is no experiment of the sort originally hypothesized, then a juror must construct the likelihood from what is known about this witness and the circumstances of the witness's observation of the cab. These conditions can include the considerations Bergman & Moore identify.[24] If the witness is more likely to err by reporting "green" (or something else besides blue—an event that I now designate as G) when the cab is blue than by reporting "blue" when the cab is green, then $p(WG|B) > p(WB|G)$. This contrasts with the previous experiment thought to establish that $p(WG|B) = p(WB|G) = .2$. By way of illustration, let us now assume that Bergman & Moore's arguments convince the juror that while $p(WG|B)$ is still .2, $p(WB|G)$ is smaller, say, .1. The pertinent ratio remains $p(WB|B)/p(WB|G)$ and it relates only to the accuracy of the witness at bar, but its value is now $(1-.2)/.1 = 8$. The posterior odds become

$$\frac{p(B|WB)}{p(G|WB)} = \frac{p(WB|B)}{p(WB|G)} \frac{p(B)}{p(G)} = 8\frac{15}{85} = \frac{24}{17} \qquad \text{(B.7)}$$

and the posterior probability is $24/(24+17) = 24/41 = .59$. There is nothing "illogical"[25] in a juror adopting this number as a measure of partial belief in the proposition that the cab was blue.

C. Potter v. Schrackle

How would a Bayesian handle the fact that Shrackle knew that the truck that he was driving had expensive cabinets in it when it struck Potter?[26] Again, the Bayesian formulation is straightforward. Let S be the event that Schrackle was speeding, let $-C$ be the event that he was not speeding, let C be the event that he knew that his truck contained expensive cabinets and let $-C$ be the event that he did not know this. The odds that Schrackle was speeding are

[24] *Id.* at 617-18.

[25] *Id.* at 618 n. 103.

[26] Potter v. Schrackle is a hypothetical discussed in Bergman & Moore, *id.* at 602.

$$\frac{p(S|C)}{p(S|-C)} = \frac{p(C|S)}{p(C|-S)} \frac{p(S)}{p(-S)}. \tag{C.1}$$

The prior odds can be obtained by envisioning all the evidence in plaintiff's story except for the stipulated event C. In other words, the Bayesian asks, if Potter and Schrackle had told the same stories without the bit about the cabinets, what would be my partial belief in Potter's claim that Schrackle was speeding? For concreteness, suppose this juror feels that these prior odds are 3 to 2:

$$p(S)/p(-S) = 3/2. \tag{C.2}$$

To elicit the likelihood ratio for the juror, we engage in the following dialogue with the juror:

Q. Continue to keep in mind the stories as told by Potter and Schrackle, excluding the bit about the cabinets. You have before you all the evidence except for the testimony about the cabinets. Given everything you know about Schrackle and the events on the day of the accident, is it more likely that Schrackle would drive fast if he had expensive cabinets in the truck than if he didn't?

A. It is hard for me to ignore testimony I have already heard.

Q. I know. Even though courts ask jurors to do this all the time, everyone knows these instructions do not always work. But you are an ideal juror, not a flesh-and-blood creature like me.

A. OK. I have erased my memory of the testimony about . . . what was it?

Q. You don't remember any testimony about expensive cabinets?

A. Expensive cabinets? No, were there any?

Q. Before we get to that, I want you to estimate the probability that Schrackle was speeding in light of everything you remember.

A. I believe the odds are 3 to 2.

Q. Now that I can inform you that there were some expensive cabinets in the truck, please think about the likelihood of there being such cabinets if Schrackle were speeding as opposed to Schrackle's not speeding.

A. I think it less probable that there would be such cabinets if Schrackle were speeding than if he weren't.

Q. How much less probable?

A. On the basis of my stock of knowledge, Shrackle seems a reckless guy. He might worry about the cabinets, but he might not.

Q. Can you be more precise about the relative likelihood?

A. Of course. I am an ideal, not a real juror. And I am a Bayesian. On balance, I would say that it is .9 times as likely to find expensive cabinets in the truck if Schrackle were speeding than if he were not.

Q. Are you speaking of the one Schrackle and the unique events in this case, or are you imagining the relative frequency of cabinets given speeding for some hypothetical collection of Schrackle-like drivers?

A. It does not help me to think about imaginary ensembles. I have considered what I know of other people like Schrackle, of course, and the likelihood ratio I have reported is informed, but not dictated, by my sense of how frequently such people would drive with cabinets. I am reporting my likelihoods for the events in this case under the evidence that has been presented to me evaluated in light of my knowledge of the world.

Conforming to the Bayesian ideal,[27] this juror substitutes the foregoing values for the prior odds and the likelihood ratio to conclude that the posterior odds are $.9(3/2) = 27/20$. The juror's partial belief in S conditioned on C (and everything else the juror knows) has changed from $p(S) = 3/(3+2) = .60$ to $p(S|C) = 27/(27+20) = 27/47 = .57$. Because the juror felt that the presence of the cabinets was a weak indicator of not speeding, the change in the partial beliefs pertaining to the litigated event is slight. As in the previous hypothetical cases, no frequentist interpretation of the prior probability or the likelihood ratio is necessary.

II. Partial Beliefs As Probabilities

Bergman & Moore's examples do not come close to proving the thesis that, in principle, an idealized factfinder cannot reason as Bayes' rule prescribes to arrive at a set of partial beliefs that could underlie a verdict.[28] In every case, a simple Bayesian treatment pro-

---

27 Bergman & Moore's juror says that he can't form the likelihood ratio because (1) he cannot remember the facts that would produce a story without C, (2) his "initial story, which took into account the cabinets, would influence the story" without C, (3) it would be impractical to repeat this process for every item of evidence, and (4) "if I'm strongly persuaded [of] Potter's story in this case . . . I would assign zero weight to the cabinets evidence in this particular case." Bergman & Moore, *supra* note 1, at 609-13. My hypothetical juror (1) has a better memory, (2) can clear his memory and construct a new story without knowledge of C, (3) is not asked to repeat this process for every item of evidence because the posterior probability is independent of the order in which conditionally independent evidence is received, and (4) appreciates the fact that probative evidence has weight even when one already has formed a strong belief. On the last point, see Kaye, *Quantifying Probative Value*, *supra* note 14.

28 Neither do Bergman & Moore advance any clearly defined, general theory that undermines the Bayesian analysis. At most, they intimate that the descriptive psychological phenomenon of embedding propositions in stories or schema precludes probabilistic assessments of propositions. That is, they seem to think that real jurors accept stories and believe all aspects of these stories with equal conviction. However, even this claim is doubtful. Why

duces the probabilities of the events in question.

However, the question of whether these probabilities are the best measure of partial beliefs remains. Why, after all, should a fact-finder adopt probabilities as the measure of partial beliefs? Why should the ideal juror in Bergman & Moore's *Potter v. Schrackle* hypothetical conclude that .57 is *the* number that specifies how firm the belief that Schrackle was speeding should be? If intuitive, untutored story construction produces a different degree of belief, why should the ideal juror strive to reconcile the two?

There is no simple, short and sweet answer to this question. The argument that has received the most attention in legal literature[29] is that anyone who fails to post "coherent" betting odds (corresponding to probabilities) on the full set of propositions that are partly believed is open to a Dutch book—a series of bets in which he or she is sure to lose money. I have nothing to add to these discussions, but it is worth noting that other arguments can be used to motivate the identification of probabilities with partial beliefs. The most fundamental analysis is axiomatic,[30] but less elaborate arguments may also serve to counter the impression that Bayesians are compulsive gamblers.

One such argument relies on "scoring rules." Suppose that a juror is given a long list of many of the propositions relevant to a disputed case. In *Potter v. Schrackle*, for example, the list might include the following:

(1) Schrackle was driving the truck that struck Potter.
(2) Schrackle was speeding.
(3) The collision damaged Potter's back.

Next to each of the listed propositions is a space for a number that represents, on a scale from zero to one, just how plausible the proposition seems. Suppose that 100 propositions are on the list, and let $Pl(p)$ represent the plausibility of each proposition p. If p is true, the juror gets a score of $[1 - Pl(p)]^2$ for the assessment of p's plausibility; if p is false, the score is $[0 - Pl(p)]^2$. The objective is to assign plausibilities so as to minimize this penalty score. Thus, if the juror assigned a plausibility of one to all true propositions and zero to all false

can't a real juror think that some propositions in the story that imparts meaning to an otherwise chaotic collection of sentences are more believable than others? And, might not a real juror find it helpful to assess the competing stories spun by counsel in light of the insights that can come from an explicitly probabilistic analysis? I do not purport to know the definitive answers to these empirical questions, but I see no grand theoretical basis for insisting that they must be answered in the negative.

29 *See, e.g.*, PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM (P. Tillers & E. Green eds. 1988).

30 For a survey of axiomatic foundations, see Fishburn, *supra* note 9.

propositions, the total penalty would be zero. If the juror got it all backwards, treating all true propositions as utterly implausible (zeroes) and all false propositions as completely plausible (ones), then the total penalty would equal 100. It has been shown that to minimize the penalty score, one must assign plausibilities that behave like probabilities.[31]

Another argument builds on the same idea of assigning plausibilities to all propositions p, q, r, and so on, as well as more complex propositions such as p & q, p or p, and $-(p \& q)$. It can be proven that the only way to assign the same plausibility to all logically equivalent propositions[32] is to use plausibilities that obey the rules for probabilities. In addition, the only way to maintain logical consistency when a new proposition z is added to our stock of knowledge is to readjust all the plausibilities by conditioning on z, that is, by changing $Pl(p)$, $Pl(q)$, $Pl(r)$, . . . , to $Pl(p|z)$, $Pl(q|z)$, $Pl(r|z)$, . . . , where $Pl(p|z) = Pl(p \& q)/Pl(z)$, ... This rule of conditionalization is just Bayes' rule.[33]

Arguments like these, however, are not conclusive. Maybe the ideal juror need not care about any penalty calculated from a scoring rule like the one mentioned above. Perhaps this juror should feel no need to regard a proposition as equally plausible when it is expressed in another form. These are matters that deserve further analysis. But the theology espoused by "Rabbi" Bergman and "Reverend" Moore will not prompt the ideal juror to lose faith in the rule of conditionalization discovered by the Reverend Bayes.

31 Lindley, *Scoring Rules and the Inevitability of Probability*, 50 INT'L STATISTICAL REV. 1 (1982). For example, consider the propositions p and its negation, $-p$. One familiar rule for probabilities is that the probability of p plus the probability of not-p is one: $Pr(p) + Pr(-p) = 1$. Suppose that the juror does not use plausibilities that have this property. Say, the juror asserts that $Pl(p) = .5$ and $Pl(-p) = .4$. If p is true, the score is $(.5)^2 + (.4)^2 = .41$. If p is false, then the score is $(.5)^2 + (.6)^2 = .61$. The juror can do better by picking plausibilities that sum to one. To see this, try $Pl(p) = .55$ and $Pl(-p) = .45$. The scores then are reduced to .405 and .605, respectively. D. LINDLEY, MAKING DECISIONS 34 (2d ed. 1985).

32 By logical equivalence, I mean truth functional equivalence. For example, p and $-(-p)$ are equivalent because if p is true then $-(-p)$ is true, and if p is false, then $-(-p)$ is false. In contrast, p and $(p \& q)$ are not equivalent because when p is true and q is false, one expression is true and the other false.

33 B. SKYRMS, CHOICE AND CHANCE, *supra* note 7, at 193.