Review Essay

The Bayesianism debate in legal scholarship

EPHRAIM NISSAN
*CMS, Queen Mary Court, University of Greenwich, Old Royal Naval College, 30 Park Row,
Greenwich, London SE10 9LS, UK*

Ron Allen and Mike Redmayne (eds.), *Bayesianism and Juridical Proof*, "Special Issue 1997" (unnumbered) of *The International Journal of Evidence and Proof* (Vol. 1). London: Blackstone Press Ltd., 1997. Pp. 253–360. ISBN 1-85431-708-3; ISSN 1365-7127.

Mike McConville and Roger Leng (eds.), *The International Journal of Evidence and Proof*, Vol. 1, No. 5 (Volume 1: 4 issues + special issue). London: Blackstone Press Ltd., 1997. Pp. 361–434.

It is obvious to me, *qua* author of this review essay and *qua* guest editor, that readers of the special issue you have in your hands should next turn to that other special issue, the one edited by Allen and Redmayne. It is devoted to the current debate about the adequacy of Bayes' theorem in statistical approaches to legal evidence. Two lead articles open the issue: Ron Allen's "Rationality, algorithms and judicial proof: a preliminary inquiry" – representing the skeptical stance – and Richard D. Friedman's "Answering the Bayesioskeptical challenge". To put it in the editorial's own words, the two papers lay down the contours of the debate, and "[i]t is to the issues raised by Allen and Friedman that the next eleven contributions are directed", followed then by two ripostes to the critics. It is a factual statement that (to say it again, with the editorial), "[i]ncreasingly, statistical evidence has found its way into criminal trials, especially in cases involving DNA evidence. The result is that legal practitioners now need more than passing familiarity with the insights of statistics and with the inference that statistics can legitimately give rise to in the evidence of experts". The skeptics' concern is with probing the delimitation of "legitimately", within that proposition. Couldn't it be the case that statistics is being fetishized, and that the prestige the tool is accorded is mesmerizing its way into misapportioning verdicts, or, less dramatically but no lesser reason for concern,

into supporting a correct outcome but for the wrong reason? Couldn't it be that we risk brewing the mythical "pebble soup", where you first put pebbles in the pot, and then go on adding the usual ingredients that would go anyway into a decent soup? Such interrogatives also arise if one is to epistemologically consider the potential (that we have tried to outline in this issue you have been browsing and hopefully reading now) of artificial intelligence for legal evidence: are we doing it for the sake of justice, or toward the greater glory of AI? Is it right to ascribe a computational tool with the eventual capability of pinpointing truth values in fact, investigation and then a probatory context in the courtroom? Should the specifications of tools be confined, instead, to unobtrusively making life easier for anybody involved in a professional capacity in the identification, organization, and comparison of the evidence? Going back to the specific theme of the special issue under review, "Bayesianism and Judicial proof", its editorial starts by pointing out that anyway, "[t]he day has passed when judges and lawyers could conduct discussions on such matters as standards of proof in qualitative language which paid no regard to quantitative methods of measurement" as cases involving DNA evidence show. Allen, a contribution by whom also opens the issue you have in your hands now, holds the chair in the name of Wigmore – the most prestigious name in legal evidence scholarship in 20th-century America – at the Northeastern University School of Law. The other guest editor of *Bayesianism and Juridical Proof*, Mike Redmayne, is now based at the Law Department of the London School of Economics. Allen's lead article in their own special issue begins (255) by a sobering, even somewhat intimidating caveat; to which extent Darwin's ghost (rather than Heraclitus') is the proper one to evoke I couldn't say, but it certainly is a precious *memento mori* when an approach is becoming (or striving to become) fashionable in a discipline, or at the meet of disciplines:[1]

> The landscape of the law has an interesting Darwinian quality. It is littered with dead algorithms and formalisms (henceforth 'algorithms'), but everywhere now algorithms are arising to take their place, and in some instances as with species hastening the extinction. Some of the defunct algorithms and their demise were quite spectacular. Legal science, the idea that the law like a science could be reduced to a set of necessary and sufficient conditions, is one good example [...]. Justice Owen Roberts's theory of constitutional law

---

[1] Probabilities have been brought into the study of judicial decision-making also in legal linguistics and semiology, from the coign of vantage of *systemic functional linguistics*, which differentiates functional choices within language as a meaning system (such as whether a clause expresses a negotiable projected idea or a non-negotiable embedded fact). The extent of the certainty attached to meanings is affected. "The precarious situation that arises when an issue divides a discourse community tends to concern a redistribution of social resources, as well as a redistribution of probabilities within paradigmatic systems on all levels of semiosis" – in the words of Iedema, whose discussion "focus[ses] on the linguistic contrasts between two common law judgements, in order to assess the possible paradigmatic redistributions in legal discourse" (p. 28 in Rick Iedema, "Legal ideology: the role of language in common law appellate judgments", *International Journal for the Semiotics of Law*, 8(22): pp. 21–36, 1995).

that to decide the constitutionality of the statute one merely needed to lay the law and the constitution side by side and compare them is another [. . .]. The Supreme Court's struggle with the varying levels of scrutiny under the equal protection clause is yet another. Less grandly, every field of law of which I am aware is littered with numerous, largely defunct, 'X-part tests' articulated to resolve some general category of problem only to have the tests disintegrate under the pressure of unanticipated developments. The field of evidence is no exception. Here perhaps the best example is the common law development of relevancy, [first leading to the articulation of relevancy rules, then to their demise]. Evidence has also experienced the demise of legal theorems. The best example is the various [false general] proofs that employing the civil burden of persuasion of a preponderance of the evidence will minimise or optimise errors. [. . .] If legal formalisms have a spotty history, what drives their creation? [. . .] First, algorithms work in other disciplines. The law, at least legal scholarship, has always struggled with an identity crisis. It [also requires that] its decisions should be predictable and correct, just like science. [. . .] [S]ome legal algorithms work pretty well. Pay your taxes by 15 April (or get an extension) or you're in duck soup, unless you have a really good reason not to, is a good example. [Yet it] succeed[s] just because of [its] nonalgorithmic qualifications [. . .]. The tension between algorithms and judgment may be reflected [. . .] today [in the s]ustained attention [. . .] being given to the nature of juridical proof.

One aspect of Allen's critique of Bayesianism also concerns what he perceives to have been unproductiveness in past debate, where players "seemed to talk past, rather than engage with, each other's arguments" (256). He lists several issues of the Bayesians that to the skeptic are not at issue; when it comes to point 4, "Whether there are any juridical fact-finding contexts in which Bayes' theorem might be useful" (258), then "[t]he Bayesian skeptic is not this skeptical", but relegates it to, possibly, "some extremely impoverished nonstatistical evidential settings", in contrast to "the typical juridical context involving a rich, highly complex set of interdependent pieces of evidence" (258). Another non-issue listed by Allen is whether Bayesianism "is 'useless' for other juridical objectives that understanding or improving the way fact-finders reason": denying that it "can be a useful analytical tool [. . .] would be astonishing because it would be tantamount to saying that a logical tool has no conceivable use" (258); "for who knows what problem might come down the road?". A footnote (fn. 4), dismisses the relevance of a few extant, allegedly successful "treats". The citation of one such "treat", "examin[ing] character impeachment evidence using a Bayesian model", has "many merits", but "does not bear at all" on the debate. When one turns to the unified list of references at the end of the thematic issue, one is faced (on p. 356) with a bracketed interpolation into the title of the entry referred to: "psycho-Bayesian [!?] analysis". Yet another non-issue is "[w]hether juridical decision making should be consistent with the results Bayes' theorem would provide were it feasible to implement Bayes'

theorem" (259): "the Bayesian skeptic has difficulty knowing what to make this argument. The primary problem is that Bayes' theorem cannot be implemented in a typical trial", because of complexity and also the untestability of whether outcomes are valid anyway. (Verdicts, after all, are a legal truth, not an oracle about factual truth.) Further down, part of the arguments assume use of jurors (261ff). Here is one concession Allen makes: "In one aspect, perhaps, juridical proof can be given a Bayesian interpretation. Perhaps deliberation brings forth new evidence (I think it does) [. . .]. This evidence, which is largely observations based on individual perspectives of jurors, perhaps can be processes in a Bayesian fashion, so long as it bears on previously identified theories, and was not already taken into account in the formation of the probability space or the assignment of initial probabilities" (269). Before considering possible future alternatives to Bayesianism (to him, incompatible) for subserving juridical proof, Allen still asks: "Maybe the law is crazy [. . .], and thus should be Bayesian. [. . .] Maybe, but the law certainly cannot be Bayesian in the same sense as some sciences can" (271). Anyway, he lists a few shortcomings with conventional beliefs about proof requirements (burdens) at trial. As to suggestions "what might juridical proof be if it is not Bayesian", Allen tries to delineate requirements for "plausible explanations" in terms of "such variables as coherence, consistency, completeness, uniqueness, economy and (yes) probability. What I think is occurring at trials, and should occur, is that parties identify their stories and try to support them with evidence, although not always in that order. [. . .] In a civil case, [the most plausible] hypothesis wins. In a criminal case, if there is a plausible hypothesis of guilt and none of innocence, the state wins; if there is a plausible hypothesis of innocence, the defendant wins. This suggests that the legal debates should move from Bayesianism to plausibility and explanations, a move perhaps the Bayesians would welcome" (274). This reviewer thinks both camps share such a goal; this explains why exponents of both camps have been welcoming the initiative of the *Artificial Intelligence and Law* special issue devoted to legal evidence. "Other tools will be necessary, and we may very well eventually reach the conclusion that wise human judgement over the domain of human affairs – while is an accurate characterisation of the primary desiderata of the law – in large measure involves knowing it when we see it. This is not a comfortable conclusion for those who look at the universe and see chaotic forces needing domestication. The problem is to ensure that the domestication does not do more harm than good, that the precision imposed is not misleading and pernicious precision", Allen befittingly concludes his paper (275).

The second lead paper, by the University of Michigan's Friedman, defends Bayesianism. "There are variations among Bayesioskeptics, as among Bayesians. Some Bayesioskeptics are rather uncompromising: for example, Alex Stein [. . .] has contended that subjective probability theory is 'vacuous' and lacks even heuristic value. Other Bayesioskeptics [e.g., Callen] acknowledge some limited usefulness for Bayesian methods but spend considerable energy criticising those methods and little or none using them" (276, fn. 1). To Friedman, "in litigated matters [. . .]

uncertainty is of a subjective nature", and "whatever the value of Bayesian methods as opposed to classical statistical methods in scientific inquiry, in litigation I believe that a subjectivist approach to probability is the only one that can offer any hope of assisting in the analysis of juridical proof" (277). Adopting a perspective of seeking greater expected utility, and formulating accordingly the degree of confidence as the standard of persuasion, p and d are the plaintiff and the defendant *qua* the party that receives the judgment; $P(\Pi)$ and $P(\Delta)$ respectively represent the probability that the facts are such that the plaintiff – or, instead, the defendant – is entitled to judgment; "for example, $U(p, \Delta)$ equals the social utility of a judgement for the plaintiff when the truth, if it were known, is such that the defendant should receive judgment" (277–278). "In a criminal case, the long-standing and solidly established view is that $U(p, \Delta)$, the negative utility of an incorrect judgment for the prosecution, far exceeds any of the other utilities in magnitude" (278). In civil cases, instead, $U(p, \Pi) = U(d, \Delta)$. Anyway, finding for the plaintiff will be optimal only if the fact-finder's degree of confidence is at least as great as

$$O(\Pi) = \frac{P(\Pi)}{1 - P(\Pi)} = \frac{U(d, \Delta) - U(p, \Delta)}{U(P, \Pi) - U(d, \Pi)} \tag{1}$$

Friedman admits that "the charge is sometimes made of 'Bayesian imperialism'. To this charge, I think, Bayesians should plead half-guilty" (278). He does not believe in the usefulness of alternative systems of probability, leading a fact-finder to results inconsistent with Bayesianism's. However, "just as it is generally best that the audience not see what is going on backstage", Friedman does "not believe that probability theory usually needs to be mentioned in the courtroom" (291), and that, in any case, it needs application to be careful: "if the limited role of Bayesian analysis is kept in mind, then I believe that the arguments of the Bayesioskeptics lose most of their force" (ibid.). Johan Bring, in his commentary, retorts: "if the limited role of Bayesianism as described by Friedman is kept in mind, the Bayesians' claim loses most of its force!" (292). Bayesianism, to say it with the title of one of the sections in Friedman's lead paper, is "[a]n only partially imperialistic view", and "an incomplete determinant of the standard of persuasion" (279). Whereas he does "not believe that the standard of persuasion necessarily can be expressed solely in terms of probability", nevertheless, to him, it "should play a large role in defining that standard" (ibid.). Friedman's addresses Allen's and others' concern with the so-called "problem of conjunction": in Friedman's own terms, "a plaintiff must prove several factual propositions" (279), and difficulties arise when the probability of their conjunction is calculated; moreover, "the number of elements into which a given claim is divided is essentially arbitrary" (280), and "the more elements a claim is divided into, the easier it is for the plaintiff to satisfy the burden with respect to each element" (ibid.). Friedman dismisses the problem by advocating a "cumulative approach", by which "the fact-finder should find for the plaintiff only if $P(A \& B) > 0.5$" (ibid.); conceding however that "[w]here the defendant raises an affirmative defence, a more complex instruction would be required" (ibid., fn.

9). Then, Friedman sets to explain away Allen's previously published objections to the cumulative approach, and, offering a few more objections himself, proposes to overcome them by policy-based devices (284). The standard of persuasion apart, objections concerning Bayesianism's adequacy to cope with the presumption of innocence are addressed, with Friedman arguing that obviating is feasible, for all of its being "trickier that has been acknowledged by good Bayesians" (285).

Next, Friedman turns to confront the legal narrative aspect *vis-à-vis* Bayesianism, "a false dichotomy" (286). He attributes "[t]he perception of a conflict [...] to a misconception about the role of Bayes' theorem in Bayesian reasoning. The theorem is important, but it is not all of Bayesianism" (ibid.). He denies "that fact-finders can, should or do go through [...] a serial updating of probability, given each new piece of evidence" (287), and besides, "not all pieces of evidence call for application of Bayes's theorem" (ibid.). Friedman tries to reconcile the story-telling model of fact-finders' tending to view an entire body of evidence, to a probabilistic representation. He had previously remarked (284, fn. 18) that "courts should not be mesmerized by an attempt to determine 'the single most probable story' ", as "there is no clear boundary line to a story; to make the fact-finding process coherent, a wealth of infinitesimally different story lines must be batched together" (ibid.). Computational complexity beyond the powers of human intellect is one more objection to Bayesianism's applicability to legal evidence – "a point that has been especially emphasised by Callen" (288, fn. 26) in several publications – and Friedman claims that "the argument is wide of the mark" (288): "a flexible template", Bayesianism "can take into account as much complexity as its user is able to handle" (ibid.). Athletes cope with a ball hustling through the air without being physicists (289), and likewise, to Friedman, when they think well, fact-finders "reach results that are roughly consistent" with rigorous use of Bayesianism, which anyway (and it may smack of a retreat into the ivory tower) is useful "as an analytical tool" (290).[2]

Friedman complains about criticism being leveled at the general level rather than about particular instances of application, short of which, "we will continue happily cooking away" (ibid.). "[T]he value of Bayesian analysis of evidence does *not* depend on the assumption that probability theory will be explicitly presented in court, through expert testimony, argument of counsel, or (where there is a jury)

---

[2] Michael Levin, in "A misuse of Bayes's theorem" (*Informal Logic*, 19(1): pp. 63–66 (1999)), pointed out a reasoning fallacy on quantifying probabilistically the rate of right judgment or, in particular, of properly accurate witness. "To say someone is right $n\%$ of the time does not mean that he is right $m\%$ of the time, $m \ll n$, nor does it imply that his accuracy depends on the frequency of his guesses. Yet these seemingly evident truths are flouted by an analysis of 'being right $n\%$ the time' which has attained wide currency in the pedagogical and legal literature" (ibid. p. 63). "A standard analysis of probabilistic reasoning in the legal and psychological literature implies that people commonly overestimate the reliability of witnesses. This paradoxical result arises from a misexplication of reliability. 'Witness is right $n\%$ of the time' ordinarily means $P(p/\text{Witness says } p) - n$, not its converse — but the standard analysis takes reliability as the converse", and wrongly so, Levin points out (ibid.).

instruction by the court", yet "the question of explicit presentation arises recurrently" (290), on occasion dramatically, whence the need to address it. Sometimes, a tutorial could even confuse jurors (290), but in some circumstances. such as when DNA evidence is presented, "it *might* be helpful to show with the use of Bayes' theorem just how powerful the evidence is, that is, what effect it should have in altering a juror's assessment of the probability of guilt" (291). In a recent case involving DNA evidence, "using the expert to suggest to the jurors that they employ an iterative algorithm" had them perceive it "to be unrealistically mechanical as well as utterly bewildering": "the problem of presentation is a tenacious one" (ibid.). Which, by the way, is where semiologist of law Bernard Jackson draws the line between what he calls the "semantics" of the legal narrative the court is called to find about, and its "pragmatics", that is to say, what is made out of the narrative through the filter of presentation. (See his papers: B. S. Jackson, " 'Anchored narratives' and the interface of law, psychology and semiotics", *Legal and Criminological Psychology*, 1(1), pp. 17–45 (1996); "Narrative models in legal proof", *International Journal for the Semiotics of Law*, 1(3), pp. 225–246 (1988); as well as his books: *Law*, *Fact and Narrative Coherence*, Merseyside: Deborah Charles Publ., 1988; and *Semiotics and Legal Theory*, London: Routledge & Kegan Paul, 1985.)

Of the eleven short contributions commenting on Allen's and Friedman's papers, the first one, by Johan Bring, from Uppsala University, makes much of Friedman's very cautiousness to turn it against Bayesianism: "Unfortunately, I do not believe that all scholars have the same sober view of the usefulness of Bayes' theorem as Allen and Friedman" (292). Bring criticizes Friedman's "story-telling Bayesian model" (292): "It is not clear how Friedman derives the factorisation in the equation above. [. . .] If the fact-finders have heard *all* the evidence, they have to condition it on *all* the evidence. The only exception could be when we have 'objective' evidence like DNA frequencies. However, all evidence that is open to subjective evaluation will be affected by the fact that we have heard other pieces of evidence" (293). Also, where to Friedman Bayesianism is "a flexible template" (which he applied to how complex to make treatment), to Bring such representations are all too elastic:

> Friedman does not seem to be so strict about the requirement that the prior should be assessed before hearing the evidence. He suggests that if a fact-finder is unhappy with her posterior she may go back and change, for example, her prior. So, if the result from the Bayesian calculation does not correspond with our intuitive belief we could alter components in the Bayesian calculations until the Bayesian answer corresponds with our intuitive belief! What, then, is the point of Bayes's theorem? (293).

Bring's conclusions commend a piece of Bayesianists' work that has recently been presented in book form (it is reviewed in this special issue separately, in Vern Walker's review essay): "I think the eminent work by Kadane and Schum on the *Sacco and Vanzetti* case proves that Bayes' theorem can be an interesting method of

studying evidence but that it is impossible for normal fact-finders to use such methods in practice. It took two of the world's leading experts several years to analyse this case" (295), and fact-finders lack both the time and the skills. Friedman's riposte does not take issue with all comments ("I will pick my spots" (348)), and does not mention Bring; it starts by pointing out that in his perception (unless it is not merely "wishful perception" (348)), "several leading Bayesioskeptics (Allen, Callen, Stein) acknowledge – with varying degrees of specificity and varying degrees of grudgingness – that standard probability theory can be useful as an analytical tool [. . .]" (348). "I do not contend that fact finders in litigation ordinarily might to be required or encouraged to make explicit or conscious use of Bayesian methods at trial. Even Robertson and Vignaux, perhaps the most unyielding Bayesians of all in the evidentiary debate, make clear in this symposium that they agree" (352).

Mississippi College's Craig R. Callen, in the comment next to Bring's, "doubt[s] that any of us skeptics would argue that formal analysis is *useless*", yet "we might argue that it chiefly serves to dramatise or illustrate some implications of ideas, which may themselves have no particular formal credentials" (296). Callen's contribution, titled "Computation and juridical proof", "adapt[s] an example which John Searle used to illustrate the difference between minds and computational programs", in order "[t]o show how Bayesianism fails to map many aspects of fact finding in the empirical word" (296). "Bayesianisim is a formal computational system", and "such a system cannot accurately represent the complexity of human decision making, in principle or otherwise, without considerable assistance from some other quarter, as yet unidentified by Bayesians" (ibid.). Obviously, AI research need pay close attention to such an argument, as AI, not just Bayesianism, is within the scope of Callen's critique. "Friedman talks about assessment of the *likelihood* of stories, but not about their formation or development" (297). Bayesian analysis has nothing to say about whether particular phenomena count as evidence for a particular conclusion. Such decisions, instead, must rely on our own associative memories and interpretive processes, including story formation, which Bayesian analysis does not purport to model" (ibid.). Moreover, "there is no Bayesian process for ascertaining the probative value of a particular piece of evidence, but merely a check to make sure that whatever subjective probabilities one uses are consistent" (ibid.). Also, the inferential process itself is not properly addressed (ibid.). I in turn propose that there is room here for intelligent technologies, including such that even Searle is not displeased with. After all, AI and subsymbolic computation have something to offer on those very counts.

Statistician Alicia L. Carriquirry (Ron Allen's interlocutor in their dialogated article in the thematic Vol. 31 (1997) of the *Israel Law Review* I am reviewing separately in the suite of book reviews of mine, now in press in *Information and Communications Technology Law*, focuses on her differences with Allen – in his "brilliantly argued work" (299) – on complexity, ambiguity and subjectiveness, on the conjunction problem, as well as on the unknown decision space problem. It is "when Allen calls for other algorithms that may fill in the alleged gaps left open by

the Bayesian paradigm" (I am using Carriquirry's own words on p. 299) that our present readership in Al & Law is directly concerned and prodded to contribute. For the unknown decision space problem, Carriquirry proposes a formal Bayesian procedure ("So perhaps" we are spared the "need to go off on a wild chase of alternative algorithms after all" (303)) as an attempt to remedy for Allen's concern about the impossibility, for fact finders, to come up with prior probabilities untainted by the presentation of the evidence.

Another statistician, Peter Donnelly from the University of Oxford, starts his discussion with "such mixed evidence cases", where "one substantive part of the evidence is statistical, while the rest is purely qualitative" (304). "Can such cases be accommodated in theories based on 'plausible explanations' or story telling? In the absence of an acceptable alternative, implementational difficulties alone cannot undermine the Bayesian approach" (305). He then "argue[s] that the use of approximations and/or bounding arguments can substantially reduce the complexity of the Bayesian approach" (304): to Donnelly, the difficulties of calculation are beyond the point, as *Bayesian fact-finders are not required to evaluate their posterior probability* [Donnelly emphasis.] All that is necessary from a computational point of view is to decide whether that probability is above or below a certain threshold" (306). Donnelly recognizes the merit of Allen's warning that "implementation of the Bayesian paradigm is in some respects more complicated that some Bayesian enthusiasts suggest" (307); nevertheless, he tersely dismisses some concerns, such as about the requirement that jurors should only assess evidence after all the evidence has been presented, or about reassessment in the light of new explanations – either to make evidence compatible with a particular hypothesis, or as new hypotheses emerge – or, then, about strong necessary conditions for application of Bayes' theorem. Donnelly distinguishes "between situations in which Bayes' theorem *can* be applied, and those in which it can be shown that under certain conditions an individual *must* act as if he were a subjective Bayesian. Not surprisingly, the conditions needed to ensure the latter *are* strong" (ibid., n. 11).

Likewise within the Bayesian camp, Carnegie Mellon University's Stephen E. Fienberg "clearly side[s] with the position articulated by Friedman", "agreeing with many (although not all) of the points he makes" (309), while disagreeing with Allen, even though some issues "require careful examination" (ibid.). Bayesian reasoning is Fienberg's ideal to which to strive:

> Sometimes we demand that a theory provide an overarching framework totally consistent with both rational and irrational legal rules and practices. Other times, we use such a theory to explain or gain insight into a particular line of legal reasoning [...] or to illustrate the irrationality of a legal ruling [...]. Yet other times we look to the theory to be useful in practical ways. Most legal evidence scholars fail to make a clear distinction between these normative, descriptive and prescriptive modes of reasoning. [310:] Allen and others rightfully begin with the descriptive and note that, when left to their own devices, few humans reason in a manner consistent with Bayes' theorem. This

is also true of real and mock juries, although they don't necessarily deviate in as dramatic a fashion as most commentators suggest [...]. But does the failure of individuals to reason informally in accordance with Bayes' theorem mean that we should throw up our hands and try to model the incoherence of everyday thinking, or rather that we should discover enough about the failure so that we can educate individuals and groups to reason more consistently and coherently? (309–310)

Fienberg disagrees with Friedman from a more extreme Bayesian stance:

Friedman counters this claim [of Allen on exceeding complexity] in several useful ways, although in the end he backs away from a full probabilistic framework with the hope that there will be some simpler way to present complex evidence. Based on the cases in which I have been involved as an expert witness [...], I believe this is simply wishful thinking. The complexity of statistical evidence will not go away, and if scholars of the theory of evidence [renounce statistics], they must come forward with a defensible alternative, one with a sound mathematical foundation that also provides sensible interpretations of real data in actual legal settings (312)

– whereas "recent fads [...] simply have not yet made the grade" (312). (In his riposte to Fienberg, Allen maintains: "I offered the Bayesian enthusiasts like Fienberg an opportunity to demonstrate how to blend trials and subjective Bayesianism, an opportunity he declined" (344).)

Let us admit, here, that the enterprise of AI for legal evidence (regardless of the Bayesianism debate) will eventually have to meet high standards, whatever the task or sets of tasks it will settle upon as being appropriate. On the spur of the good will it has been meeting on the part of legal evidence scholars, it will be essential that AI as being applied to the domain will not only pursue the development of building blocks or either integrated or disparate tasks, but also keep in touch with specific concerns of the users it ultimately seeks for itself. Applications such as Carmelo Asaro's DAEDALUS tool, now adopted by the judiciary in Italy and described in a paper to appear, have such a delimited scope and niche of use as to make them uncontroversial. Not so anything that will have to do with reaching a verdict; a tool for the prosecutor is not the same as a tool for affecting and orienting the outcome of the fact-finders' decision making in court.

In his contribution to the special issue under review, D. H. Kaye chooses to "focus on a simple issue – the burdens of persuasion" (313). In Kaye's terms, Friedman's formulation "is a call to consider second-order probabilities – degrees of confidence in a first-order judgment" (313); whereas one decade earlier, Kaye had suggested that first-order probabilities could do, now he is willing to watch others as they try to respond to the new challenge. On the other hand, Kaye is critical of Allen's criticism. "Allen, it seems, would prefer a rule that minimises the actual frequency of errors (or that produces the particular mix of errors of one kind as opposed to another). Of course, that is neither an attack on 'algorithms' nor

a demand for greater generality, but a call for a different criterion for choosing a decision rule" (314). Kaye's line of reasoning is that this actually vindicates the role of probability and statistics. "Only in the context of such an analytical framework can one distinguish between the concepts like utility, probability and error that must be examined to arrive at a suitable conception of the burdens of persuasion" (315). Kaye dismisses the relevance of juror psychology to the specific topic, albeit recognizing its interest in its own right; having discussed it elsewhere, he merely provides citations to his previous work, as "repetition is not persuasion" (315). The most recent reference is to Kaye's paper of 1991, "Credal probability" (*Cardozo Law Review*, 13: pp. 647–656)).

The next commentator is Richard Lempert, in "Of flutes, oboes and the *as if* world of evidence law" (316–120). "Allen is a prominent Bayesioskeptic while I have been labelled a Bayesian enthusiast" (316) – which Lempert, resurrecting an old parable about the cold war, likens to the conflict of a flute and an oboe who nevertheless, once summer is over and the orchestra reassembles in the fall, have come to resemble each other and are now clarinets. Lempert opens by making the mischievously tendentious, farfetched claim that Allen has rather come over to the Bayesian camp (somehow foreshadowing the parting "Parthian arrow" in Friedman's riposte). Legal scholars are never short of forensic rhetoric, it seems. (Lempert is both Professor of Law and Professor of Sociology at the University of Michigan, and describes himself as "a social scientist who has long been interested in the empirical aspects of trial fact finding" (320).) After a list of points coaxed into construal as pro-Bayesian concessions on the part of Allen, Lempert concedes one instance of the reverse: "Finally, I agree with Allen that except in a few situations involving statistical evidence, fact finders should not be instructed in the use of Bayes' theorem or told to apply Bayes' theorem to non-statistical legal evidence" (317). He also agrees about problems being posed by complexity. "But this does not move me towards flutehood because I have never taken a contrary position" (ibid.). Lempert acknowledges that jurors are neither intuitive Bayesians, nor to be urged to be formal Bayesians.

Lempert, by way of a riposte, engages Allen's reference to the former, then goes on to develop a discussion on what he identifies as being the enabling condition for the points of agreement between the two scholars, made "possible because much evidence law, like the uses of Bayes' theorem that Allen criticises, lives in an *as if* world", that is, "*as if* the frailties of human existence and the substantive complexities of actual trial evidence did not exist. Thus, evidence law often presumes that its fact finders have abilities that exceed the capabilities of most humans. Allen has his eye on the actual world of trials. I, and to a large extent Friedman, are writing for the more abstract *as if* world of much evidence law" (318). "For example, jurors are presumed to use information about an accused thief's past convictions to decide if the accused might be prone to lie but not as evidence that she is prone to steal, though research suggests jurors cannot so limit the impact of what they hear" (ibid.). Moreover, "in the law's *as if* world, judges can be trusted to ignore improper

evidence even if they have not explicitly recognised its inadmissibility"(ibid.). Also problematic in the real world, though not in the *as if* world, is the ascription to trial jurors and judges of the ability to distinguish truthful and accurate witnesses from lying or inaccurate witnesses even though body language and other demeanour cues are often ambiguous or misleading" (318–319). "Perhaps more astonishingly", the *as if* world ascribes fact finders with the ability of making sense in appropriate ways of "the conflicting testimony of honest scientists, all of whom have needed years of advanced study to master their fields to the degree where the law will allow them to speak" (319). Assuming that fact finders' "thought processes are appropriately modelled by Bayes' theorem" for some purposes is also not unconceivable in an ideal world (ibid.). In Lempert's own reading of the U.S. Federal Rules of Evidence 401 and 402, he claims that the law itself ascribes such unreal assumptions to fact finders.

> It is no accident that evidence law has created an unreal world in which to operate. Some scholars would see this as inescapable. [...] But it is not ethereal theory that has created this condition. Practically speaking a judge's analytic tasks [...] are considerably eased by the acceptability of [idealizing juror behavior]. Indeed, since people behave differently and their behaviour is often poorly understood, the law can find itself in trouble when it proceeds on an image of behaviour designed to approximate actual behaviour rather than some idealised vision of how rational people act. Thus, where courts attempt to make law that realistically appraises human behaviour, their initial decisions are often followed by ill-disguised retreats that lead to bodies of precedent that are neither behaviourally nor analytically consistent (319).

Lempert provides examples from American law. He then concludes by proposing that if the goal of Allen's lead article is "to challenge evidence law's *as if* world through greater understanding of how legal fact finders process information and make decisions, he is embarking on and is inviting others to join him in a truly radical challenge to received law and, indeed, to ways of thinking about law. In doing so he points to exciting paths for evidence scholars to explore" (320). Whereas already H. Munstenberg, *On the Witness Stand: Essays on Psychology and Crime* (New York: McClure, 1908), proposed such a direction for evidence scholarship – Lempert points out – "[w]hat is noteworthy is for as keen an analytic lawyer as Allen to move his attention to the empirical" (ibid.), including moving the attacks on Bayesian law from the analytic to "the case against Bayesianism he makes in his current article" (ibid., n. 11). Allen's riposte does not specifically object to Lempert and his reading of the lead article.

Andrew Ligertwood, from the University of Adelaide, comments on Friedman from a Bayesioskeptic viewpoint. Bayesian analysis, to Ligertwood, "creates a false sense of logic and certainty" (324). He rejects counterclaims "that this is just a matter of education" (ibid.), and that it's up "to the parties to decide which hypotheses and issues they wish to run", with Bayesian analysis applying to the resulting information. "This all has to me an appearance of pragmatism, the justification

of which the Bayesians cannot provide. Therefore even if we accept the canons of mathematical probability these just cannot take us very far in the process of reconstruction. We need another logic, a broader logic to keep us on the rails. And it must be in this broader logic that the secret of proof lies, not in a mere Bayesian analysis. This is the root of my Bayesioskepticism" (ibid.). Ligertwood does not accept Friedman's solution for Allen's point about the problem of conjunction, as well as to the Bayesian counterclaim about how to treat initial probabilities (325).

To Albert Madansky, "[s]ince a juror is a person, not a computer, the use of probabilities and data in litigation should (with the exclusion of computational errors) conform to what people 'do' with probabilities and data, not with what normative theory says they 'ought to do'. The lawyer's duty is to provide correct and complete probabilistic computations of the likelihood ratio (and not allow the juror to perform the computation himself). As to the posterior odds ratio, I would rather have the juror make that determination himself" (326), even though it will not he consistent with Bayesianism. Madansky's discussion is mathematical.

John A. Michon, from Leiden, titled his contribution: "The time has come to put this debate aside and move on to other matters" (331), which he then describes as being "a direct quote from Allen's article. I agree wholeheartedly with this conclusion – and so does Friedman, it seems. Evidently the issue is ultimately a nonproblem. It derives from mixing up two distinct levels of scientific explanation, known as the *normative* level and the *descriptive* level, respectively" (ibid.). This sounds like being in line with Lempert's stance. In the normative camp, "signal detection theory has evolved into a generic normative model of decision making under uncertainty, which Michon and F. J. Pakes have applied to law"; for this reference, see their "Judicial decision-making: a theoretical perspective" (in R. Bull and D. Carson, eds., *Handbook of Pychology in Legal Contexts*, Chichester: Wiley, 1995, pp. 509–525; the book is reviewed separately in this special issue). Michon in Allen and Redmayne's special issue goes on: "No material system will ever meet the strict requirements of a normative theory. This is the basic insight expressed by Herbert Simon's principle of *bounded* rationality" (332). "In sum, no normative theories are descriptive, psychological theories. Neither is Bayesian theory" (ibid.). The title of Michon's contribution in the same forum – which reminds me of philosopher Marcelo Dascal's paper "The beyond enterprise", on the use of programmatic 'beyond' in scholarship, especially in titles (in J. Stewart, ed. *Beyond the Symbol Model*, Albany, NY: SUNY Press, 1996, pp. 303–314) – "should be taken as an invitation to gain a better understanding of the cognitive make-up of fact finders" (Allen and Redmayne, p. 333). "The legal enterprise being a complex work environment it should be accessible to the ergonomic approach that has been applied successfully to a great many organisations of comparable complexity" (ibid.). Yes, but could they ever [be let to] condemn the innocent? "[N]orms and values have not found a place in ergonomic analysis" (ibid.). Michon wishes to enrich the latter with empirical investigation in such a norm-and-value driven system. (Having mentioned Dascal's work, note that in the context of law

and the normative-descriptive distinction, there is some substantively relevant work of his, namely, his papers: "Epistemología, controversias y pragmática", *Isegoría* 12, pp. 8–43 (1995), and "La balanza de la razón", in: Oscar Nudler, ed., *La Racionalidad: Su Poder y sus Limites*, Segundo Coloquio Bariloche de Filosofía, 1994, "La racionalidad en debate", Buenos Aires: PAIDOS, pp. 363–381 (1996).)

Bernard Robertson and G. A. (Tony) Vignaux – whose recent, readable and important book is reviewed separately in this special issue – answer from a staunchly Bayesian position, one by one, the twelve points that Allen had raised on p. 263ff. Robertson and Vignaux claim they found those twelve points useful inasmuch as they prod into clarification, but they also warn against "spending so much effort on fundamental (and often the least tractable) questions that we forget to move forward, to use Bayesian reasoning to solve new problems and to improve its use as a yardstick against which to measure judges' statements about how facts should be reasoned" (335). They concede that Bayesian reasoning is generally not for jurors to use consciously. "What is asked is that normative statements about reasoning should be consistent with it" (336). "There is no requirement that jurors come to the same factual conclusions or the same overall probability assessments, let alone assign the same weight to individual items of evidence [. . .]. In fact Bayesian reasoning offers a framework for identifying the reasons for disagreement, which may lead to some of the protagonists changing their minds" (337). Given "that reality is unknowable and untestable" and that often some evidence is withheld from the jury, "[t]he most that can be asked is that they make optimum use of the available information. Bayesian reasoning provides the tools for doing that" (ibid.). In point 7, Allen had stated: "The question facing jurors is not 'guilty or not guilty'; it is these facts or those facts. Once that decision is made, application of the legal categories typically follows deductively" (266). Robertson and Vignaux retort that they had, earlier on, themselves "made the point that 'innocence' or 'not guilt' [sic] is not a useful hypothesis": "Facts need to be considered in the light of well-formulated hypotheses. What the appropriate alternative hypotheses might be will sometimes not be apparent until late in the trial. But this is a problem for trial procedure rather that for Bayesian reasoning. It is noticeable that witnesses such as scientists and doctors tend to press for evidence to be exchanged prior to trial so that the alternative hypotheses can be properly considered" (336). Researchers in Al & Law ought to pay attention especially to point 11 in Robertson and Vignaux:

> It is true that Bayesian reasoning is only a method of reasoning about hypotheses in the light of evidence. It does not formally provide a process for generating new hypotheses although in practice its application can stimulate consideration of new hypotheses. Hypothesis generation is the least understood and most complex aspect of the whole problem. But this is not a criticism of Bayesian reasoning until and unless a theory of hypothesis generation appears which produces results incompatible with the Bayesian model (337).

Hypothesis generation would be a very ambitious goal indeed, in the long run, for AI models of legal evidence.

Jerusalem-based legal scholar Alex Stein opens by distinguishing between normative and descriptive uses of Bayesian analysis of juridical proof, and claims that "[i]n practice, there are virtually no examples of 'trial by mathematics' [...]. Juridical Bayesianism therefore has no explanatory power" (339), which does not amount to "a wholesale denial of any utility of probability theory in law" (ibid.). Exceptionally, it still can apply to determination of facts, but Stein acknowledges the value of probabilistic analysis especially "as a heuristic device" which "(among other methods) may further our understanding of important evidentiary doctrines, such as those regulating burdens of proof, relevancy, hearsay, expert testimony and character evidence", and "usefully explain, and perhaps even predict, the operation and the overall utility of the examined doctrines" (ibid.), yet at a descriptive macro-level, "about longruns, either directly or focusing upon typical or average cases" (ibid.). However, does juridical Bayesianism – having "a manifestly limited explanatory power" – nevertheless " have any normative appeal?" (ibid.). Stein challenges (340) Friedman's argument in favor, for all of Friedman's moderation *vis-à-vis* Robertson and Vignaux within the Bayesian camp. "The logic of Friedman's system strikes me as impeccable. His discussion, however, does not specify the epistemic standards that should be fed into this system" (ibid.), and Stein contends that if the assumption is "that the standards regarded as good for non-Bayesian fact finding, as presently conducted by judges and jurors, will be good enough for his system as well", then this is questionable (ibid.). Stein's Bayesioskepticism had already found place in Vol.1 of *The International Journal of Evidence and Proof* (on pp. 25–47).

Stein's comment on Friedman is followed by Allen's and then Friedman's conclusive articles. Allen responds to Fienberg, then to Donnelly and Robertson and Vignaux (reproaching all of these for unsatisfactorily dealing with the use of subjective probabilities), then to Carriquirry and finally to Kaye. The rebuttal is followed by Friedman's own response, which is organized around themes, rather than per commentator. The special issue certainly does not reduce the gap between the two camps, but it displays a rich-textured debate which does progress indeed in putting in evidence some topics. It is required reading for researchers from AI & Law willing to devote efforts to modelling evidence, whether quantitative modelling of probativity is involved, or, perhaps more usefully, not so or only marginally so (or then, for example, in helping a prosecutor to construct his or her strategy, instead of trying to grope for reconstructing the legal narrative for the purposes of affecting the fact finding).

In the other, regular journal issue under review (Vol. 1, No. 5), Rosemary Pattenden, of the University of East Anglia in Norwich, discusses a judge's recommendation, recently made, "that a civil court should have a discretion to exclude relevant and otherwise admissible evidence" in order "to avoid civil courts being troubled by superfluous testimony" (361). Should England follow the U.S. (or Australia) in this regard? and how? The paper contains an interesting discussion of relevance. Ed Cape, from the University of the West of England, is concerned with

the right to silence, and "the problem of how to advise a client where explaining the reasons for advice would inevitably involve disclosure of privilege information, and therefore amount to waiver" (400). Cope concludes that custodial legal advice – advice being given at the police station by defence lawyers – has had its value diminished in light of recent court decisions. The third paper in the issue, by Roger Leng and Richard Taylor, is also from England: "Living with the discretionary admission of written hearsay in the Crown Court". It is followed by a case note by Susan Nash (from London), on the disclosure of journalists' sources, and then by a "Noticeboard". For all of the British slant clearly shown by this regular issue, the thematic issue demonstrates the attraction of the new journal for scholars from the American tradition. Its disciplinary area clearly warrants the sustained attention of legal evidence scholarship worldwide. Note that unlike some law journals, *The International Journal of Evidence and Proof* is refereed. It will clearly make as much progress as it will manage to attract at international authorship. The special issue, reviewed here in such detail, testifies to the forum's achievement and potential.[3]

---

[3] I would like to refer the interested reader to a review essay by Mike Redmayne, "A likely story!" (*Oxford Journal of Legal Studies*, 19: pp. 659–672 (1999)), discussing William A. Dembski's book *The Design Inference: Eliminating Chance Through Small Probabilities* (Cambridge University Press, 1998), "a detailed analysis of a process by which we can distinguish real coincidences [. . .] from patterns that can only be attributed to design", Redmayne's goal being "to suggest that Dembski's work is important to evidence scholars by showing how it sheds light on certain evidentiary problems" (660). On probability and utility in a civil case, see Redmayne's paper "Standards of proof in civil litigation", *The Modern Law Review*, 62: pp. 167–195 (1999). In that paper, he "defended the 0.5 'more probable than not' standard as the default standard for civil cases", while conceding that his arguments are "open to the criticism that they attempt to impose precision and clarity on an area where these are not virtues" (194), and moreover to a critique to the effect that "[t]here are sometimes good reasons for conceptual ambiguity in the common law. By not committing themselves too deeply to any particular theory, courts narrow the scope for disagreement among judges and allow themselves some leeway to respond to the merits of individual cases" (195), a counterargument which Redmayne rejects for civil standards of proof.