

# Notre Dame Philosophical Reviews

---

2003.11.10

Author 

Richard Swinburne (ed.)

## Bayes's Theorem

Published: November 10, 2003

Swinburne, Richard (ed.), *Bayes's Theorem*, Oxford University Press, 2002, 160pp, \$24.95 (hbk), ISBN 0197262678.

Reviewed by **Branden Fitelson**, University of California'Berkeley

---

This is a high quality, concise collection of articles on the foundations of probability and statistics. Its editor, Richard Swinburne, has collected five papers by contemporary leaders in the field, written a pretty thorough and even-handed introductory essay, and placed a very clean and accessible version of Reverend Thomas Bayes's famous essay ("An Essay Towards the Solving a Problem in the Doctrine of Chances") at the end, as an Appendix (with a brief historical introduction by the noted statistician G.A. Barnard). I will briefly discuss each of the five papers in the volume, with an emphasis on certain issues arising from the use of probability as a tool for thinking about *evidence*.

In the first essay, Elliott Sober contrasts Bayesian accounts of evidential support with an alternative, non-Bayesian, likelihood-based approach. The crux of Sober's non-Bayesian proposal involves the following sort of claim about *contrastive* evidential support:

---

Evidence  $E$  favors hypothesis  $H_1$  over hypothesis  $H_2$ .

---

Here, the alternative hypotheses  $H_1$  and  $H_2$  need not be mutually exclusive. Sober proposes that we should unpack this relational concept of favoring using *likelihoods*, as follows:

---

Evidence  $E$  favors hypothesis  $H_1$  over hypothesis  $H_2$  if  $\Pr(E | H_1) > \Pr(E | H_2)$ .

---

This principle is sometimes called the "Law of Likelihood" (see Royall (1997) for the history and theoretical basis of this "law"). From a Bayesian (and, I think, *intuitive* point of view), this "law" is far from obvious. Consider a case in which  $E$  entails  $H_1$  but fails to entail  $H_2$ . Intuitively, in such a case,  $E$  should *favor*  $H_1$  over  $H_2$ . After all,  $E$  *guarantees* the truth of  $H_1$ , but fails to guarantee the truth of  $H_2$ . It is important to note that the "Law of Likelihood" is *inconsistent* with this intuitive principle. That is, there can be cases in which  $\Pr(E | H_1) > \Pr(E | H_2)$ , *despite* the fact that  $E$  entails  $H_2$  but fails to entail  $H_1$ . Of course, these will be cases in which  $H_1$  and  $H_2$  are *not* mutually exclusive, but a *likelihoodist* cannot object to such counterexamples on *these* grounds (since mutual exclusivity is *not* a requirement for the likelihoodist's "favoring" relation). A proper, Bayesian theory of contrastive confirmation, on the other hand, need not have this undesirable consequence.

Bayesians typically understand relational support in terms of *non-contrastive confirmation*. For a Bayesian,  $E$  supports (or confirms)  $H$  – in a *non-contrastive* sense – just in case  $E$  raises the probability of  $H$  (on a suitable, rational credence function). There have been various proposals concerning how a Bayesian ought to measure the *degree* to which  $E$  confirms  $H$ , or  $c(H, E)$ , for short (see Fitelson (1998) for a survey). But, no matter which Bayesian  $c$ -measure one favors, one would be inclined to define *contrastive* support (or *favoring*) in terms of this *non-contrastive* confirmation measure  $c$ , as follows:

---

Evidence  $E$  favors hypothesis  $H_1$  over hypothesis  $H_2$  if  $c(H_1, E) > c(H_2, E)$ .

---

Interestingly, this “reduction” of contrastive support to non-contrastive (Bayesian) confirmation need not be at odds with the “Law of Likelihood”. As it turns out, there is one (and only one, out of all the historical proposals!) Bayesian measure of confirmation that *entails* the “Law of Likelihood”, assuming this standard Bayesian definition of favoring in terms of confirmation (see Milne (1996)). This is important, as it shows that the Bayesian need not reject the “Law of Likelihood.” However, those who think the “law” is false (like myself) would be forced either to abandon the reductive principle stated above, or to choose a different measure of non-contrastive confirmation. Indeed, many have opted for the latter approach. While I would recommend endorsing *both* the former approach *and* the latter approach, it is worth mentioning that the following *weakened* version of the “Law of Likelihood” should be acceptable to *all* parties here, Bayesian or otherwise:

---

Evidence  $E$  favors hypothesis  $H_1$  over hypothesis  $H_2$  if

---

---

$\Pr(E \mid H_1) > \Pr(E \mid H_2)$  and  $\Pr(E \mid \neg H_1) \leq \Pr(E \mid \neg H_2)$ .

---

Joyce (2003) shows that this principle is satisfied by *all* reductive Bayesian confirmation-theoretic approaches to favoring (that is, *all* Bayesian measures of confirmation  $c$  will lead to definitions of “favoring” that satisfy this weak likelihood principle). This is a nice way to see precisely where Bayesian and non-Bayesian accounts of evidential support come apart. Bayesians are perfectly happy to talk about the likelihoods of the *denials* of alternative hypotheses:  $\Pr(E \mid \neg H_1)$  and  $\Pr(E \mid \neg H_2)$ . But, non-Bayesian Likelihoodists will not feel comfortable with such probabilities, since they involve *averaging* over the likelihoods of concrete alternative hypotheses. And, the “weights” in these averages will depend on the dreaded *prior probabilities* of the alternative hypotheses:  $\Pr(H_1)$  and  $\Pr(H_2)$ . While Bayesians are happy to use priors in their account of evidential support, non-Bayesians like Sober are strongly opposed to such a move, since they think the prior probabilities are (in general) subjective and that they lack probative force. Ultimately, it seems to me, whether terms like  $\Pr(E \mid \neg H_1)$  should be countenanced in our theory of evidence will depend on the overall relative adequacy of Bayesian *vs* non-Bayesian accounts of evidential support. Howson (pp. 52–53) argues in his contribution to this volume that such likelihoods are *crucial* for properly understanding evidential support. And, I am inclined to agree (see Fitelson (2001) for some further reasons why). Indeed, even *non*-Bayesians will use such terms *sometimes* – when it seems to be *essential* to obtaining the right answers about *contrastive* evidential support (see Royall (1997, pages 1–2), and even Sober (2003) for some clear examples of this kind).

Sober’s paper concludes with a discussion of recent *instrumentalist*, non-Bayesian approaches to statistical inference. Here, he highlights the work of the Japanese statistician Akaike, which aims to show how the *simplicity* of a model can be tied to its *predictive accuracy*. This is a very important area of research in contemporary statistics and also in the philosophy of science. Sober argues that Bayesian approaches to these issues and problems cannot adequately account for the importance of simplicity as a factor in determining how predictively accurate a statistical model is. Howson, and other Bayesians, are usually not convinced by such arguments. And, in fairness to the Bayesian approaches, I think there is more that can be said on this score (for a nice Bayesian discussion of simplicity in this context, see Rosenkrantz (1977)). I conclude my discussion of Sober’s paper with a detail that the minute reader may find puzzling. In the first part of his paper, Sober talks about “favoring,” which, presumably, involves evidence favoring the *truth* of one hypothesis over another (not, say, favoring the *predictive accuracy* of one over another), but in the second part he talks only about comparative judgments of *predictive accuracy* and *not* about truth. It is unclear to me how the likelihoods appearing in Akaike’s theorem are to be interpreted. Are they still capturing what the evidence says about the *truth* of competing theories, or are they merely containing information relevant to assessing

relative *predictive accuracy*? It is interesting that (either way) likelihoods would then seem to be essential *both* to the instrumentalist *and* to the non-instrumentalist (who is concerned with evidence regarding the *truth* of competing theories). It would be nice to know how and why likelihoods are able to play this dual role.

In contrast to Sober's contribution to this volume, the papers of Howson, Dawid, and Earman adopt a Bayesian stance. The first part of Howson's paper contains a wealth of historical, philosophical, and statistical wisdom. He discusses the role of Bayesian methods (and, by contrast, some of their most notable non-Bayesian rivals) in statistical theory and practice, beginning with the very first Bayesian methods used by Laplace (and Bayes himself), leading all the way up to the most recent foundational issues addressed by Bayesian statisticians and philosophers, including debates about "informationless" prior probabilities, and the importance of simplicity in hypothesis (or model) choice. Howson's treatments of Fisherian and Neyman-Pearsonian statistical methods (and philosophies) are particularly informative and useful (the analogies with Popperian and hypothetico-deductive conceptions should be especially illuminating for philosophers). And, Howson's discussion of Lindley's Paradox is refreshing (it seems to me that not enough philosophical ink has been spilt over this important statistical conundrum).

The second part of Howson's paper (on which I will dwell a bit) is written from a more "logical" point of view. Here, he proposes a systematic, and general Bayesian (non-deductive, of course) "logic," which is described in a way that makes it sound strongly analogous to (classical) *deductive* logic. He talks about "consistency" and "soundness" and "completeness", *etc.* Some of the logicians among us will probably have deep worries about this analogy, and no doubt they will view use of this logical terminology as a non-trivial stretch. I must confess, I found myself feeling rather uncomfortable about the degree of force with which Howson pushes the analogy. I will focus here on Howson's notion of "inconsistency," but I think similar worries will apply to his other "logical" notions. When a (classical) logician talks about *inconsistency*, it is a notion that is *directly* relevant not only to decision-making and other (broadly) pragmatic disciplines, but also to *epistemology* (understood here in a traditional, *non-pragmatic* sense). It's not entirely clear to me that Howson's notion of "consistency" has such direct relevance to epistemology. Here, I am *not* worrying about the problems involving prior probabilities mentioned above. I am willing to *grant* (*arguendo*) that *they do* have epistemic and probative force. What I do not see is why someone who is "inconsistent" in Howson's sense should feel any *epistemic* pressure to revise their degrees of belief. It seems logically consistent with Howson's "inconsistency" that such an agent's degrees of belief are *inter alia* as accurate as they have ever been (or ever will be). This is (arguably) *not* the case when the agent's beliefs are *logically* inconsistent. In that case, the agent *knows* there is something *false* in what they believe (and this is transparently a *bad thing*, from an epistemic point of view). What is the analogous thing that an "inconsistent" agent (in Howson's sense) *knows* that would inspire them to change their degrees of belief? In this connection, it seems to me that Howson's discussion is somewhat vague. He presents his "logic" without crucial details concerning the proofs of the key theorems that purport to forge the strong analogy between deductive logic and his Bayesian "logic". For instance, Howson does not explain how the additivity axiom follows from his "consistency" assumptions (indeed, he even claims to establish the "inconsistency" of violations of *countable* additivity, which is *even more* controversial). This leaves one wondering whether the compelling objections to Dutch Book arguments that have been voiced by philosophers like Schick (1980) and Maher (1993) might have some bearing on Howson's approach. Such philosophers seem to provide examples of cases in which it seems perfectly *rational* to violate the additivity axiom (and, therefore, Howson's "consistency"). It would be nice to hear Howson explain what, precisely, makes such agent's degrees of belief "bad" or "irrational" (in *any* compelling sense). More traditional logicians may want to have a look at Carnap's (1950) insightful discussion of the relationship between deductive and inductive logic. Carnap's inductive logic program may have failed, but its aim was to provide a notion of partial entailment that was *logical* in the *very same sense* (not merely in an *analogous* sense) that deductive logical consequence is *logical*, and thereby to *avoid* a pragmatic and/or subjective turn in inductive logic (which seems implicit – although now deeply buried – in Howson's talk of "betting quotients" and "fairness"). It seems to me that this aim may still be achievable (albeit, probably in a non-Carnapian way), and until it is demonstrated that this goal cannot be achieved, perhaps it would make more sense to reserve the term "logic" for the non-pragmatic, non-contingent, and objective conception that traditional logicians have in mind. In the meantime, why not just stick with the term "rational", as opposed to "logical" when characterizing Bayesian accounts of credence? Would anything really be lost?

Dawid's paper provides a very clear, simple, and sound introduction to the use of Bayesian theories of evidential support (and weighing evidence) in legal contexts. A fair amount of work has been done in this area over the past thirty years or so, and Dawid's paper serves as a nice overview of the basic techniques that are applied by Bayesians in the context of legal evidence. One of the best things Dawid does is to make very clear the distinction between prior probabilities (degrees of *belief*) and likelihood ratios (degrees of *support* or *weight of evidence*). Many of the same issues discussed above in connection with Bayesian theories of evidential support arise in concrete and simple examples in Dawid's paper. Dawid proposes the likelihood ratio measure  $l(H, E) = \Pr(E | H) / \Pr(E | \neg H)$  as the proper Bayesian measure of degree of support. This measure has been skillfully defended by I.J. Good for many years (see Good (1985)), and more recently has been shown to have various advantages over other Bayesian measures of confirmation (see Eells and Fitelson (2000), and Fitelson (2001)). Importantly, because of its sensitivity to the "catch-all" likelihood  $\Pr(E | \neg H)$ ,  $l$  violates the strong "Law of Likelihood" discussed above (endorsed by Sober). And, yet, as Dawid's examples illustrate, it often seems *crucial* to take account of such terms in our assessments (both contrastive and non-contrastive) of weight of evidence. In this sense, Dawid's legal examples provide a nice testbed for clashing intuitions in the Bayes/non-Bayes controversy about evidential support. I think Dawid's examples provide further reasons to worry about the legitimacy of the strong "Law of Likelihood," and further reasons to retreat to Joyce's (2003) *Weak Law of Likelihood*.

Earman's paper can be viewed as a sampler of a much longer essay he has written [Earman (2000)] on Hume's arguments concerning miracles (an essay which I highly recommend, by the way). Earman provides a detailed historical trace of the arguments of Hume and his contemporaries concerning the possibility of compelling testimony about the occurrence of miracles. By carefully and skillfully applying Bayesian techniques to these arguments, Earman ends up with some very interesting (albeit somewhat anachronistic) new reconstructions of these infamous historical arguments. By and large, Earman's reconstructions are accurate and novel, and his analyses are trenchant. His Bayesian treatment of multiple testimonial reports is especially illuminating. The only complaint I have about this paper is that it may focus too heavily on *posterior probabilities*  $\Pr(H | E)$  of the various hypotheses  $H$  in question, given the various sorts of evidence  $E$  he considers. It would also be interesting to see parallel analyses done which focus more on the *likelihood ratios*  $l(H, E)$  that result in each of the reconstructions. I suspect the ensuing facts about *degree of support* would be harmonious with Earman's conclusions about *degrees of belief* in these cases. But, examining things from the *weight of evidence* perspective (as Dawid does in the legal context) may shed further light on some of the issues and arguments. This is a minor complaint, and Earman is to be commended for the rich historical/philosophical tale he tells, and for the interesting applications of Bayesian machinery he musters.

The final contemporary paper in this collection (aside from Swinburne's solid introductory piece on which I have chosen not to comment explicitly) is Miller's brief (but important) essay on the propensity interpretation of probability. Roughly, the propensity theory recommends interpreting  $\Pr(X | Y)$  as the (presumably, *causal*) *propensity*  $Y$  has for bringing about  $X$  (usually, in some experimental context). Popper (1957) was one of the first to endorse a propensity interpretation of conditional probability, and many others have followed suit since. Humphreys (1985) pointed out that there seem to be deep problems with the existence and interpretation of the "inverse propensity"  $\Pr(Y | X)$ , since (presumably)  $Y$ 's having a causal propensity to bring about  $X$  does *not* imply  $X$ 's having a causal propensity to bring about  $Y$ . But, if "Pr" is to satisfy the probability axioms, then it must also satisfy *Bayes's Theorem*, which would imply a perfectly well-defined and interpretable *inverse probability*  $\Pr(Y | X)$ . This became known as Humphreys's Paradox. Many people came to believe that Humphreys had shown that propensities *cannot* satisfy the axioms of probability (or Bayes's Theorem). [Indeed, it seems that some people already believed this before Humphreys' paper appeared (see Fetzer and Nute (1980)).] In his contribution to the volume, David Miller shows that this is not the case. Indeed, Miller sketches a perfectly coherent and sensible propensity theory that is also a *probability theory*. As such, Miller shows how to diffuse Humphreys's Paradox and restore the satisfaction of Bayes's Theorem for propensities. As it turns out, there are various ways to mitigate Humphreys' paradox in this sense. See Gillies (2001) for extended discussion of several approaches, including Miller's.

The volume closes with an Appendix containing a very polished reproduction of Bayes's classic "An Essay Towards the Solving a Problem in the Doctrine of Chances". The Essay still reads very well, and it should be on every probabilist's "must read" list. I feel quite comfortable saying something almost as glowing about this entire volume. I found this book very edifying and clear,

and the debates and issues it encompasses are of great importance for contemporary philosophy of probability, statistics, and decision-making. I highly recommend this book to anyone with interests in these areas, and I commend Swinburne for putting together this neat little book.

## References

- Carnap, R., 1950, *Logical Foundations of Probability*, Chicago: University of Chicago Press.
- Earman, J., 2000, *Hume's Abject Failure - The Argument Against Miracles*, Oxford: Oxford University Press.
- Eells, E. and Fitelson, B., 2002, "Symmetries and Asymmetries in Evidential Support," *Philosophical Studies* 107: 129–142.
- Fetzer, J. and Nute, D., 1980, "A Probabilistic Causal Calculus: Conflicting Conceptions," *Synthese* 44: 241–246.
- Fitelson, B., 1999, "The plurality of Bayesian measures of confirmation and the problem of measure sensitivity." *Philosophy of Science* 66: S362–S378.
- Fitelson, B., 2001, "A Bayesian Account of Independent Evidence with Applications," *Philosophy of Science* 68: S123–S140.
- Gillies, D., 2000, "Varieties of Propensity", *British Journal for the Philosophy of Science* 51: 807–835.
- Good, I. (1985). "Weight of evidence: A brief survey," In *Bayesian Statistics, 2* (Valencia, 1983), pp. 249–269. Amsterdam: North-Holland.
- Humphreys, P., 1985, "Why propensities cannot be probabilities," *The Philosophical Review* 94: 557–570.
- Joyce, J., 2003, "Bayes' Theorem", *The Stanford Encyclopedia of Philosophy* (Fall 2003 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/bayes-theorem/>
- Maher, P., 1993, *Betting on Theories*. Cambridge: Cambridge University Press.
- Milne, P., 1996, "Log[p(h/eb)/p(h/b)] is the one true measure of confirmation," *Philosophy of Science* 63: 21–26.
- Popper, K., 1957, "The propensity interpretation of the calculus of probability, and the Quantum Theory," in S. Körner (ed.): *Observation and Interpretation in the Philosophy of Physics*.
- Rosenkrantz, R., 1977, *Inference, Method and Decision*. Dordrecht: D. Reidel.
- Royall, R., 1997, *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.
- Schick, F., 1986, "Dutch Bookies and Money Pumps," *Journal of Philosophy* 83: 112–119.
- Sober, E., 2003, "Likelihood and the Duhem/Quine Problem," unpublished manuscript.