

Notre Dame Philosophical Reviews

2005.10.15

Author 

Richard Jeffrey

Subjective Probability: The Real Thing

Published: October 15, 2005

Richard Jeffrey, *Subjective Probability: The Real Thing*, Cambridge University Press, 2004, 140 pp, \$21.99 (pbk), ISBN 0521536685.

Reviewed by Branden Fitelson, University of California, Berkeley

Richard Jeffrey was one of the all-time greats in formal epistemology, and this was his last book. In classic Jeffrey style, what we have here is a short, dense, and incredibly rich and engaging monograph. It is simply amazing how much wisdom is packed into this little book.

Before getting down to Serious Bayesian Business, Jeffrey begins with an extended acknowledgements section, which contains a heartfelt, emotional, and informatively autobiographical letter of farewell and thanks. The letter is addressed to "Comrades and Fellow Travelers in the Struggle for Bayesianism", and its author is introduced to the reader as "a fond foolish old fart dying of a surfeit of Pall Malls". As someone who only barely knew Dick Jeffrey (but hopes to be a Comrade in the aforementioned Struggle when he grows up), I was deeply touched and inspired by this introductory section of the book. It's no wonder that he was so beloved and respected both as a philosopher and as a man.

The first chapter provides an excellent introduction to the basic concepts of subjective probability theory. Both the formal probability calculus as well as its interpretation in terms of betting quotients for rational agents (the main application discussed in the book) are clearly and concisely presented here. This includes very accessible and clear explanations of "Dutch Book" arguments, conditional probability, and Bayes's Theorem. There are many useful exercises, and (as always) plenty of wise remarks and notes along the way. Jeffrey's style is highly effective pedagogically, because he tends to introduce things using snappy examples. Only after whetting the reader's appetite with such examples does Jeffrey invite the reader to think more systematically and theoretically. As such, this chapter would be a suitable (maybe even ideal) way to start an advanced undergraduate course on probability and induction (or inductive logic). Indeed, I plan to try it myself the next time I teach such a course.

Chapter two explains how subjective probability can be used to provide an account of the *confirmation* of scientific theories. The basic idea is to model inductive learning (typically, involving observation) as an event (called an *update*) that takes the agent from an *old* subjective probability assignment to a *new* one. If this learning process leads to a greater probability of a hypothesis (H) -- i.e., if $new(H) > old(H)$ -- then H is said to have been *confirmed* (presumably, by whatever was learned during the update). Here, Jeffrey uses examples from the history of science to frame the discussion. Historical illustrations of both the Duhem-Quine problem and the problem of old evidence are treated here (I will return to Jeffrey's discussion of the problem of old evidence later in this review). In keeping with Jeffrey's pedagogical style, no precise theory of updating is developed at this stage (although some hints and puzzles are presented, which naturally lead the reader into wondering how such a theory might go). At this point, we just see some basic concepts applied to some simple historical examples. Precise theories of probabilistic update are discussed in the next chapter. From a pedagogical point of view, I suggest thinking of chapters two and three as operating together (I suspect that some students might have trouble following the details of the accounts exemplified in chapter two without delving into some of the more theoretical material in chapter three along the way).

In chapter three we get a masterful primer on the two main Bayesian theories of learning (probabilistic update). The classical theory of *conditionalization* (in which learning is modeled as *conditionalizing* on a proposition explicitly contained in the agent's doxastic space), as well as Jeffrey's more general theory of *probability kinematics* (in which learning is modeled as an event that alters an agent's credence function, but not necessarily by explicit conditionalization on a proposition) are compared

and contrasted in a very illuminating way. We also get a pithy presentation of Jeffrey's "radical probabilist" epistemology, which was the philosophical motivation for his generalization of classical Bayesian conditionalization. There are two main reasons why Jeffrey saw a need to generalize classical conditionalization. First, classical conditionalization assumes that all learning is learning *with certainty*, since, whenever we conditionalize on a proposition E , we must subsequently assign E probability 1. Second, classical conditionalization presupposes that there is always a *statement* (in the agent's mentalese) that expresses the precise content of what was learned during an update. Jeffrey conditionalization weakens both of these assumptions, thereby providing a more general (and more "radically probabilistic") framework for learning. The theoretical and philosophical aspects of this framework are laid out in chapter three.

Before moving on to chapters four and five (which have to do with foundations and applications of subjective probability in statistics), I would like to digress with a few critical remarks on Jeffrey's account of the problem of old evidence presented in chapter two. The problem of old evidence is a problem (first articulated by Clark Glymour) for the traditional Bayesian theory of confirmation which takes *conditionalization* as its learning rule. According to this classical approach, $new(H) = old(H | E)$, and E confirms H iff $old(H | E) > old(H)$. Hence, once E is learned, it cannot confirm any hypothesis *thereafter*, since all *subsequent* probability functions will have to assign probability 1 to E [i.e., $new(E) = 1$, and so $new(X | E) = new(X)$ for all X , and no *subsequent* confirmation of *any* X by E is possible].

But, intuitively, there seem to be cases in which we *do* want to say that E confirms H *even though we have already learned* E . For instance, Einstein knew about (E) the anomalous advance of the perihelion of Mercury, many years before he formulated his theory of general relativity (H) which predicts it. Nonetheless, it seems reasonable for Einstein to have judged that E confirms H (in 1915) when he learned that H predicts E . But a classical Bayesian theory of confirmation *cannot* undergird his claim. [Many Bayesians respond to this problem by saying that, while Einstein's *actual* credence function in 1915 did not undergird the desired confirmation claim, some *historical* or *counterfactual* credence function does (e.g., the credence function he would have had, if he had not learned about the perihelion data). I will not discuss such approaches here.] Dan Garber provided a clever alternative explanation of confirmational judgments in such cases. Garber suggested that, while E did not confirm H for Einstein in 1915, the fact that H entails E (which Einstein *did* learn in 1915) did. The idea here is to model Einstein as an agent who is *not logically omniscient*. Garber does this by adding a new statement to our (sentential) probability model of Einstein's epistemic state. This statement gets *extrasystematically* interpreted as " H entails E ". Garber then assumes that Einstein has *some* knowledge about this entailment relation (that if X is true and " X entails Y " is true, then Y must also be true), but he does not know whether or not " H entails E " is true. Then, one can give constraints (historically plausible ones, even) on Einstein's credence function which ensure that " H entails E " confirms H in the classical Bayesian sense.

Jeffrey speaks approvingly about this Garberian approach to "logical learning" and old evidence in chapter two. But he then goes on to sketch an alternative account based on Jeffrey conditionalization. On Jeffrey's account (which is rather tersely presented in chapter two), we assume that there are two learning events: the *empirical* update in which E is learned, and the *logical* update in which " H entails E " is learned. Jeffrey places various constraints on these two updates so as to ensure that, at the end of the two updates, H has a greater probability than it did before the two updates. Thus, H is confirmed by the combination of the empirical and logical updates. There are lots of moving parts and assumptions in Jeffrey's account (it's considerably more complex than Garber's conditionalization approach). I won't get into these details here (although I think some of these assumptions are rather worrisome). Rather, I'd like to focus on the *motivation* for a Jeffrey-conditionalization approach in the first place (in light of Garber's elegant, pre-existing classical conditionalization approach). Recall the two motivations (in general) for abandoning strict conditionalization in favor of Jeffrey conditionalization: (1) that sometimes learning is *not* learning with certainty, and (2) sometimes there is no statement in the agent's mentalese that expresses what was learned. The first motivation (1) cannot be relevant here, since (a) E *must* be learned with certainty in order for the problem of old evidence to get off the ground (if E is *not* learned with certainty, then E *can* still confirm H in the classical Bayesian sense, and there is *no* problem -- this is why *even Jeffrey* models the empirical update as a *strict* conditionalization), and (b) there is no reason to suppose that " H entails E " is *not* learned with certainty here (and even if there were, it is unclear how that would help to resolve the problem anyway). So, whatever Jeffrey sees as lacking in Garber's approach, it must have something to do with (2).

But Jeffrey concedes that E is expressed by a statement in the agent's (sentential) mentalese (namely, " E "). So, it seems that the *only* motivation for using Jeffrey conditionalization rather than strict conditionalization to model logical learning (and to use this logical learning to account for the old evidence problem à la Garber) is the worry that " H entails E " is not expressed by any statement in the agent's mentalese. Indeed, Jeffrey seems to presuppose this in his account sketched in chapter two. I don't find this a very compelling worry. After all, Garber has shown how to use *extrasystematic interpretation* of one of the sentences of the agent's language to model an agent's learning " H entails E ". One might respond on behalf of Jeffrey by complaining that having a sentence which is *extrasystematically* interpreted as " H entails E " is not the same thing as having a statement that *systematically* expresses " H entails E ". That's true, but I don't see why it's a problem for Garber's approach. It is quite common in the context of classical Bayesian confirmation theory to extrasystematically interpret statements in a sentential language as having first-order logical content which outstrips their systematic (propositional-logical) content. For instance, in Bayesian approaches to the ravens paradox, (atomic) sentences in simple languages are extrasystematically interpreted as monadic first-order claims like "All ravens are black", and some of the (extrasystematic!) logical implications of these extrasystematic interpretations are *crucial* for proving the requisite "theorems" about the probability models in question. So, unless there is some reason to think that such applications of classical Bayesian confirmation theory (which trace back to the origins of the discipline) need to be re-worked Jeffrey-style, so as to avoid the use of such "extrasystematic interpretations", I don't see why Garber's approach needs to be re-worked Jeffrey-style either. That said, I think Jeffrey's approach to old evidence and logical learning is both novel and clever. I just wonder whether its extra complexity and assumptions are really warranted, in light of Garber's simpler, classical approach.

Chapter four contains a perspicuous and sophisticated introduction to the concept of *expectation*, and its relation to *probability*. Both unconditional and conditional expectation are expertly (and accessibly) covered here, along with their (sometimes subtle) connections to unconditional and conditional probability. This is something we (unfortunately) rarely see in a book on the philosophy of subjective probability. But, it is essential for a thorough understanding of the foundations of the subject (especially as they were developed by de Finetti and others in the 20th century). In particular, the basics of expectation are prerequisite for grasping a key concept discussed in chapter five: *exchangeability*. Exchangeability is considered by many to be the single most important concept in the foundations of subjective probability. But it is almost never discussed in introductory texts on probability and inductive logic (at least, those that philosophers are likely to read). In chapter five, Jeffrey provides a survey of some of the central results involving the concept of exchangeability. The most important of these are various forms of and variations on de Finetti's *representation theorem* for subjective probability, which provides a key to unlocking the mystery of how subjective probabilities can be obtained (non-capriciously) by updating on statistical information. This is some of the most technically (and philosophically) challenging material in the book. But this chapter (especially) repays a careful work-through. I would say that the material in this chapter will be most challenging for students (even those with some background in probability). I would also say that those interested in the relationship between subjective and objective probability (*e.g.*, probability in statistical mechanics) will find this chapter very illuminating and thought provoking (many references to excellent related work in statistics and physics are included here). Those who want a deep understanding of the foundations of subjective probability and its relationship to contemporary statistical science would be well served by a careful study of chapters four and five of *The Real Thing*.

Chapter six (the final chapter of the book) is all about Jeffrey-style rational decision theory. Here, the reader will find a very effective crash course on the basics of the theory of rational decision first outlined in Jeffrey's classic essay *The Logic of Decision*. The presentation here benefits from many years of reflection since the publication of *The Logic*. In the very final section of the book (to my mind, one of the most interesting and sophisticated sections therein), we hear a completely new take from Jeffrey on the Newcomb problem. The Newcomb problem has plagued decision theorists (especially those of Jeffrey's ilk) for over thirty-five years. Here, at the very end of his very last work, Jeffrey renounces much of what he had been saying about that thorny problem for many years. In the process, he provides many wonderful new insights and ideas. This is the mark of a great philosophical mind (or, in his words, "a fond foolish old fart"). Even the last pages of his last book involve radical re-workings of age-old resolutions of the deepest philosophical puzzles. Richard Jeffrey was one of the greatest philosophers of probability, induction, and rational decision we have known. His last book has given me a healthy dose of his wisdom. May it do the same for you.
