What is Wrong with Strict Bayesianism?

Patrick Maher

University of Illinois at Urbana-Champaign

1. Introduction

Bayesian decision theory in its classical formulation supposes that for any rational agent and for any possible state *x* of the world, there is a number *P*(*x*) which represents the agent's judgment of the probability of *x*. Similarly, the theory assumes that for any possible outcome *y* of the agent's actions, there is a number *u*(*y*) which represents the utility or value of *y* to the agent. Given these assumptions, the theory is able to define the *expected value* for the agent of any act. Bayesian decision theory then recommends that the agent should choose, from amongst the available acts, one which has maximal expected utility.[1]

The assumption that an agent has a determinate personal probability function has often been attacked. Two criticisms have become familiar. One criticism is based on a factual claim: it is said that real agents do not in fact have a determinate probability function, and further that they lack the computational resources to have such a probability function. The other criticism is based on a normative claim: it is said that agents often lack sufficient evidence to warrant them in adopting any particular probability function, and under these circumstances it would be irrational for them to adopt a determinate probability function (even if they were capable of doing so).

These criticisms have led many decision theorists to think that Bayesian decision theory should be generalized to cover cases in which the agent does not have determinate probability judgments. One way of achieving this generalization which has been widely endorsed in recent years is to suppose that the agent's probability judgments can be represented by a (non-empty) *set* of probability functions, rather than a unique function. Advocates of this approach include I.J. Good (1952, p. 11), Richard Jeffrey (1985), Isaac Levi (1974, 1980), Brian Skyrms (1984, p. 29), and Bas van Fraassen (1985, p. 249). In this paper, I will refer to theories of this kind (i.e., theories which allow that an agent's probability judgments are representable by a set of probability functions) as *weak Bayesianism*. By contrast, the classical form of Bayesianism, which requires that an agent have a determinate probability function, will be called *strict Bayesianism*.[2]

PSA 1986, Volume 1, pp. 450-457

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193145 Published online by Cambridge University Press

Weak Bayesians suppose that their theory is superior to strict Bayesianism in just those respects in which they deem strict Bayesianism to be deficient. Thus those weak Bayesians who see the problem with strict Bayesianism as being that real agents don't or can't have precise probability judgments suppose that real agents do or can have indeterminate probability judgments representable by a set of probability functions. And those weak Bayesians who see the problem with strict Bayesianism as being that precise probability judgments are often unjustified suppose that weak Bayesianism does not require agents to do anything which is unjustified. In this paper, I will critically examine both these putative advantages of weak Bayesianism, and show that neither is an advantage weak Bayesianism can properly claim.

2. Applicability

Let us begin with the claim that real agents lack the computational resources to have a determinate probability function. One argument for this view has been advanced by Gilbert Harman, who reasons as follows:

> suppose that probabilities are to be assigned to *N* unrelated atomic propositions and their various truth functional combinations. In the general case, this requires making at least $2^N - 1$ explicit assignments (for example to all but one of the strongest conjunctions including each atomic proposition or its negation). Given a modest 300 unrelated atomic propositions, more than 10^{90} explicit assignments would be needed. This is a very large number. It has been estimated, for example, that there are about 10^{78} atoms in the entire observable universe. (Harman 1980, pp. 154 f.)

Harman concludes that "it will not be feasible for an intelligent creature to operate in terms of subjective probabilities." But Harman here overlooks the fact that probabilities can be assigned to large classes of events in a tractable way using a general function or schema. Examples of such general representation abound in mathematical statistics. For example, a statistician might suppose that the error in a certain measurement has a standard normal distribution; this means that for any (measurable) set *A* of real numbers, the probability that the difference between the real value and the measured value is in *A* is equal to

$${}_A(1/\sqrt{2\pi})e^{-x^2/2}dx.$$

Since *A* might be any one of an uncountable number of sets, this formula assigns probabilities to uncountably many events. Thus contrary to Harman, finite beings can and do assign probabilities to more events than there are atoms in the observable universe.

Another argument which is supposed to show that a finite agent could not possibly have a determinate probability function has been put forward by Jeffrey, who claims that "the probability of a single proposition requires a capacity to store no end of digits if the probability might be just any real number in the unit interval." (1985, p.117). Jeffrey claims that the solution to this putative problem is to allow agents to have indeterminate probability judgments.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193145 Published online by Cambridge University Press

Jeffrey is right, of course, that only an infinite being could name every number in the unit interval. But this does not show that finite beings cannot name some numbers in the unit interval, and use these names to express precise probability judgments. For example, if someone says that the probability of a particular coin landing heads is exactly 1/2, then they have assigned a precise probability to this event. Furthermore, once we recognize that individual names are not the only way of referring to numbers, we can see that finite beings like ourselves not only can, but in fact do, succeed in referring to <u>every</u> number in the unit interval, and are able to assign every one of these numbers as the probability of some event. The example of the normal distribution mentioned above provides a demonstration of this: someone who holds that the error in a measurement has a normal distribution has in fact assigned every number in the unit interval as the probability of some event or other. And this sort of thing is done all the time.

Thus the arguments which are supposed to show that real agents <u>could not possibly</u> have the determinate probability function required by strict Bayesianism are fallacious. But do real agents <u>in fact</u> have determinate probability functions? A number of weak Bayesians have urged that they do not (e.g., Good 1976, p. 25), and in this they would seem to be correct. For example, most agents are surely not committed to any precise value for the probability of rain tomorrow. However, this fact is not a reason for preferring weak Bayesianism to strict Bayesianism; for as I shall now argue, real agents also do not have indeterminate probability judgments, in the sense required by weak Bayesianism.

Weak Bayesianism replaces the strict Bayesian requirement of a determinate probability function with the requirement that there be a determinate <u>set</u> of probability functions which represents the agent's probability judgments. However, the same sort of considerations which suggest that agents do not have precise probabilities for many events also indicate that agents are in many cases not committed to any definite set of probability functions. Subjects show just as much discomfort in specifying precise bounds for their indeterminate probability judgments as they do when asked to specify a precise probability.

Good (1962, p. 81) has proposed to deal with this situation by supposing that an agent's indeterminate probability judgment is representable, not simply by a set of probability functions, but rather by a probability distribution over these probability functions, i.e., by a "higher order" probability function. However, the existence of a determinate higher order probability function would imply the existence of a determinate lower order probability function, namely the expected value of the lower order probability functions. To avoid this result, Good assumes an unending sequence of probabilities of higher and higher order. But one can hardly suppose that real agents have an infinite sequence of higher order probabilities. If anything, the supposition that real agents have an infinite hierarchy of higher order probability functions would seem to be even less realistic than the supposition that they have a determinate probability function, or are committed to some definite set of probability functions.

Thus the fact that real agents do not have precise probability judgments for some events is not a disadvantage of strict Bayesianism vis-à-vis weak Bayesianism. And we have also seen that the arguments which purport to show that finite agents could not possibly have a precise probability function are fallacious. Thus there is no basis for saying that weak Bayesianism is more applicable to real agents than strict Bayesianism. Let us turn, then, to the claim that weak Bayesianism is superior to strict Bayesianism as a normative ideal.

3. Justification

Suppose you are told that an urn contains a number of balls, each of which is either black or white; you are, however, given no information about the proportion of black or white balls. A ball is about to be drawn from the urn. Let $\underline{W}$ be the event that the ball drawn is white. According to strict Bayesianism, there should be some number $\underline{P}(\underline{W})$ which represents your judgment of the probability of $\underline{W}$. But writers such as Fisher, Neyman, Pearson, Kyburg and Levi have urged that in cases such as this no definite probability judgment can be justified by the available evidence. They conclude that strict Bayesianism is incorrect as a normative ideal, since it requires the agent to do something which is unjustifiable.

Levi (1974, 1980) claims that weak Bayesianism is superior to strict Bayesianism in this regard. We shall see, however, that this claim is also incorrect. Anyone who thinks it is wrong to satisfy the requirements of strict Bayesianism ought also to think it is wrong to satisfy the requirements of weak Bayesianism, and for essentially the same reason. The argument to this conclusion requires us to examine the application of weak Bayesianism to decision making.

Once we move from strict to weak Bayesianism, the expected utility of an act ceases to be well defined. Hence weak Bayesianism needs to replace the strict Bayesian principle of maximizing expected utility with some other principle of rational choice. One proposal for such a replacement has been made by Good (1952, p. 9). Letting P be the set of probability functions which represents the agent's probability judgments, Good's proposal is that the agent may rationally choose, from amongst the available acts, one which maximizes expected utility relative to any one of the probability functions in P.

Note that Good's proposal does not prevent the agent from choosing, in each decision problem, an act which maximizes expected utility relative to the same probability function. It also does not prevent the agent from deciding in advance to follow the strategy of always choosing an act which maximizes expected utility relative to one of the probability functions in P. But to decide to always choose acts which maximize expected utility relative to some probability function is the same thing as adopting that probability function as one's subjective probability. Hence Good's proposal permits the agent to adopt a determinate probability function. In fact, according to Good's proposal, the adoption of a determinate probability function is just as rational as any other way of proceeding. So anyone who thinks that the adoption of a determinate probability function is sometimes irrational will also have to regard Good's version of weak Bayesianism as permitting irrationality.

A weak Bayesian who disapproves of determinate probability functions in some contexts will thus want to have a different decision rule to

Good's. Such a rule has been proposed by Levi. According to Levi's favored criterion for rational choice, the fact that an act has maximal expected utility relative to some probability function in **P** is a necessary but not sufficient condition for that act to be a rational choice. A further condition which needs to be added, according to Levi, is that amongst those acts which maximize expected utility relative to some probability function in P, the chosen act should maximize the minimum utility. Thus Levi supplements Good's criterion for rational choice with the maximin criterion.[3]

Levi's criterion for rational choice is, unlike Good's, incompatible with strict Bayesianism. For although in each decision problem the acts which are admissible by Levi's criterion are also admissible relative to one of the probability functions in P, the probability functions for which this is true will not be the same in all decision problems; and hence Levi's criterion does not permit the agent to decide to always maximize expected utility relative to one of the probability functions in **P**. But to be inconsistent with strict Bayesianism is one thing, and to be superior to it is something else. So we need to ask: does Levi's criterion for rational choice avoid the alleged normative defect of strict Bayesianism, i.e., are agents who follow Levi's decision rule any more justified in what they do than ones who act as strict Bayesians?

We are imagining a situation in which an agent is justified in ruling out any probability function not in some set **P**, but is not justified in preferring any one of the functions in **P** to any other. Under these conditions, both Levi and strict Bayesianism agree that the agent should always choose an act which satisfies Good's criterion, i.e., an act which maximizes expected utility relative to some probability function in **P**. Levi and strict Bayesianism also agree that a further condition needs to be added to Good's criterion; they disagree only about what that further condition should be. According to strict Bayesianism, the condition which needs to be added to Good's criterion is that the same probability function in P should be used for all decision problems. And according to Levi, the condition which needs to be added to Good's criterion is that the chosen act should have at least as great a minimum utility as any of the acts counted as admissible by Good. Thus a necessary condition for Levi's criterion to be normatively superior to strict Bayesianism is that there be some justification for applying maximin in addition to Good's criterion.

Now there are circumstances in which the use of the maximin criterion would seem to be justified. These are circumstances in which one is playing a zero-sum game against a clever opponent. But this is a very special case. Furthermore, in this case strict Bayesianism will give the same recommendation as maximin.[4] But what justification is there for using maximin in other contexts? There are infinitely many alternative rules to maximin. For example, one could maximize the maximum utility, rather than the minimum. Or one could maximize the mean (or some other combination) of the maximum and minimum utilities. Or one could maximize the mean of all the possible utilities of an act. And so on. Or one could choose a probability distribution over the states, and maximize expected utility relative to that distribution. What justifies the selection of maximin in preference to these other rules?

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193145 Published online by Cambridge University Press

There is no satisfactory answer to this question, either in Levi's corpus or elsewhere in the literature; no justification for the general use of maximin has ever been produced. This being the case, it seems reasonable to conclude that there is in fact no justification for the general use of maximin. Consequently, agents who act in accordance with the maximin criterion are doing something which is unjustified; they have adopted an unjustified decision rule. Thus Levi's normative criticism of strict Bayesianism is equally applicable to his own version of weak Bayesianism.

4. Conclusion

We have now shown that all the usual claims made for the superiority of weak Bayesianism over strict Bayesianism are fallacious. Real agents can have a determinate probability distribution. The vague probability judgments of real agents are no more representable by a set of probability functions than they are by a unique probability function. And if it is irrational to adopt a determinate probability function, then that is an objection to Good's version of weak Bayesianism as well as to strict Bayesianism, and a parallel objection applies to Levi's version of weak Bayesianism. So if there is a good reason for abandoning strict Bayesianism in favor of weak Bayesianism, it isn't any of the reasons usually cited. Weak Bayesians owe us a new explanation of what defects in strict Bayesianism their theory corrects.

I anticipate one response to this challenge, namely that weak Bayesianism is a more _general_ theory than strict Bayesianism, and is superior to it on that account. For while the arguments in this paper indicate that weak Bayesians are in no position to say it is normatively _incorrect_ for an agent to satisfy the requirements of strict Bayesianism, we have given no argument to think that agents are rationally _obliged_ to satisfy these requirements either, and under these circumstances it would not be implausible to think that the more general theory should be preferred. However, if generality is all that makes weak Bayesianism superior to strict Bayesianism, then Levi's decision rule will have to be abandoned in favor of Good's. For with Levi's decision rule, weak Bayesianism is _not_ more general than strict Bayesianism, as we observed above. Thus the question raised in this paper--the question of whether and how weak Bayesianism is superior to strict Bayesianism--has, at the very least, implications for the form weak Bayesianism should take.

Notes

[1] This recommendation is subject to the proviso that the states are suitably independent of the acts.

[2] Weak Bayesians have usually allowed utilities as well as probabilities to be indeterminate. In this paper I concentrate on the probabilities, though much of the discussion would apply equally to utilities.

[3] Levi also has another condition on rational choice (what he calls _P_-admissibility), but this is vacuous in virtually all practical decision problems. Levi also toys with the idea of using leximin in place of

maximin, which would mean that acts with the same minimum utility would be ranked according to their second-worst outcomes, and so on. Neither of these refinements makes any difference to the issues being discussed in this paper.

[4]For this we need to supplement strict Bayesianism with an account of probability dynamics, such as that given by Skyrms (1984, pp. 75-80).

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193145 Published online by Cambridge University Press

References

Achinstein, P. and Hannaway, O. (eds.). (1985). Observation, Experiment, and Hypothesis in Modern Physical Science. Cambridge, MA: MIT Press.

Churchland, P.M. and Hooker, C.A. (eds.). (1985). Images of Science. Chicago: University of Chicago Press.

Good, I.J. (1952). "Rational Decisions." Journal of the Royal Statistical Society B 14: 107-114. (As reprinted in Good (1983). Pages 3-14.)

---------. (1962). "Subjective probability as the measure of a non-measurable set." In Logic, Methodology, and Philosophy of Science. Edited by E. Nagel et al. Stanford: Stanford University Press. Pages 319-329. (As reprinted in Good (1983). Pages 73-82.)

---------. (1976). "The Bayesian influence, or how to sweep subjectivism under the carpet." In Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science. Volume 2. Edited by C.A. Hooker and W. Harper. Dordrecht: Reidel. Pages 125-174. (As reprinted in Good (1983). Pages 25-55.)

---------. (1983). Good Thinking. Minneapolis: University of Minnesota Press.

Harman, G. (1980). "Reasoning and Explanatory Coherence." American Philosophical Quarterly 17: 151-157.

Jeffrey, R.C. (1985). "Probability and the Art of Judgment." In Achinstein and Hannaway (1985). Pages 95-126.

Levi, I. (1974). "On Indeterminate Probabilities." Journal of Philosophy 71: 391-418.

-------. (1980). The Enterprise of Knowledge. Cambridge, MA: MIT Press.

Skyrms, B. (1984). Pragmatics and Empiricism. New Haven: Yale University Press.

van Fraassen, B.C. (1985). "Empiricism in the Philosophy of Science." In Churchland and Hooker (1985). Pages 245-308.