Philosophy Compass 6/5 (2011): 321–332, 10.1111/j.1747-9991.2011.00398.x

Bayesianism II: Applications and Criticisms

Kenny Easwaran*
University of Southern California

Abstract
In the first paper, I discussed the basic claims of Bayesianism (that degrees of belief are important, that they obey the axioms of probability theory, and that they are rationally updated by either standard or Jeffrey conditionalization) and the arguments that are often used to support them. In this paper, I will discuss some applications these ideas have had in confirmation theory, epistemology, and statistics, and criticisms of these applications.

1. Bayesian Confirmation Theory

One of the important early uses of probability theory in philosophy was in (Carnap 1950), which argued that one notion of probability (corresponding to something like rational degree of belief, though he had in mind a more purely logical notion that gives rise to constraints on degree of belief, rather than working with degree of belief directly (Carnap 1945)) gives the best analysis of the scientific notion of confirmation. Like Hempel before him (Hempel 1945), Carnap sought not just a qualitative theory of whether a given piece E of evidence confirms, disconfirms, or is neutral with respect to a hypothesis H, but also a theory of when E_1 confirms H_1 more strongly than E_2 confirms H_2, and possibly even an absolute quantitative scale of these degrees of confirmation. However, Carnap's book had an unfortunate ambiguity between two different measures of confirmation, pointed out in (Popper 1954). Because this ambiguity was so pervasive in Carnap's book, he never fixed it, although he acknowledged the problem in the second edition. As it is often put, the ambiguity is between an *absolute* notion of confirmation as 'firmness' ($P(H \mid E)$) and an *incremental* notion of 'increase in firmness' ($P(H \mid E) - P(H)$), either one of which can be seen as something like the amount of support E gives to H. Since this has been clarified, most Bayesian measures of confirmation have aimed at something more like the latter, 'incremental' notion, than the former, though the confusion is an easy one to make, and has often been made by others since.

Thus, Bayesian confirmation theorists traditionally analyze the qualitative notion of confirmation by saying that E confirms H iff $P(H \mid E) > P(H)$, though there is still controversy about whether the probability function is the degrees of belief of some particular agent at a time or some other probability function, and whether extra propositions beyond E should be conditionalized on. Bayes' theorem[1] states that $P(H \mid E) = P(H)P(E \mid H)/P(E)$, so that (as usual, assuming for now that none of the relevant probabilities are 0) $P(H \mid E) > P(H)$ iff $P(E \mid H) > P(E)$. Given the standard Bayesian picture of update by conditionalization, we can see that if $P(H)$ is the degree of belief a scientist had in H before conducting her experiment, and E is the unique proposition she learns as the result of performing the experiment, then $P(H \mid E)$ will be her degree of belief in H after performing the experiment. Thus, an experiment confirms a hypothesis (as expected) iff the scientist's degree of belief in the hypothesis increases as a result of performing the experiment.

There is much standard terminology in Bayesian confirmation theory. Understandably, $P(H)$ is called the 'prior', and $P(H \mid E)$ is called the 'posterior', but confusingly, $P(E \mid H)$ is called the 'likelihood' of the hypothesis – when discussing Bayesian confirmation theory, the apparent synonyms 'probability' and 'likelihood' have different meanings. Thus, E confirms H iff the posterior is greater than the prior, or equivalently iff the likelihood is greater than the initial degree of belief in the evidence.

Although Bayesians agree on all this, there is much disagreement about the measure of a *quantitative* notion of confirmation. Because of the clarity and success of the 'increase in firmness' understanding of confirmation, it has often been assumed that the appropriate measure of confirmation is $P(H \mid E) - P(H)$. However, it can easily be seen that this measure makes it impossible to confirm a theory very strongly if the prior is high, since the posterior must always be at most 1. As a result, many other measures have been proposed, disagreeing about the relative strengths of different confirmations, but they all agree about when there is confirmation, disconfirmation, and irrelevance.[2] These include $P(H \mid E) - P(H \mid \neg E)$, $P(E \mid H)/P(E \mid \neg H)$, and $P(H \mid E)/P(H)$ (which by Bayes' theorem also equals $P(E \mid H)/P(E)$) among others.[3] Each of these measures has been argued for and against in a variety of publications, and each has also been applied to the analysis of various apparent cases of confirmation.

A famous series of analyses, discussed in depth in (Fitelson 2006), concerns the 'paradox of the ravens'. This paradox arises from an observation by Hempel that his theory of confirmation predicts that the observation of a black raven (as expected) confirms the hypothesis that all ravens are black, but that the observation of a non-black non-raven will also confirm this hypothesis. A number of Bayesians eventually showed that on their account, neither of these confirmation relations necessarily holds, but that under assumptions that seem very plausibly true in the real world, the confirmation given by a black raven is much, much greater than the confirmation given by a non-black non-raven.

However, the fact that there are a variety of measures of confirmation is still troubling. As pointed out in (Fitelson 1999), several analyses of this paradox (and others) depend strongly on which particular measure of confirmation is being used. However, (Christensen 1999) argues that this variety of measures is actually a good thing, because some can capture intuitive comparisons of confirmation that the others fail to recognize, and thus he argues that Bayesian confirmation theorists ought to be pluralists about the measurement of confirmation. These measures all agree about whether or not there is confirmation in any given case, but they can often have drastic disagreements about the relative importance of different pieces of evidence.

2. *Problems for Bayesian Confirmation Theory*

Despite the successes of Bayesian confirmation theory in dealing with old paradoxes and analyzing many standard cases of confirmation, many problems for the theory have been posed.

2.1. Popper–Miller Argument

One apparently serious problem for Bayesian confirmation theory comes from (Popper and Miller 1983). Popper had already long argued that there could be no such thing as inductive support (instead endorsing a view on which deductive falsification and falsifiability were the only relevant epistemic features of a scientific theory) but this argument claimed to show that in particular, Bayesian confirmation couldn't give inductive support for a theory. They argue that since H is equivalent to $(H \vee E)\&(H \vee \neg E)$, the support E gives to H can be separated into the support E gives to $H \vee E$ and the support E gives to $H \vee \neg E$. But the first piece of the support is clearly deductive (and thus not inductive), while the second piece of support is always neutral or negative, so there is no positive inductive support.

However, this argument rests on several clear mistakes, as described in the many responses cited in (Earman 1992, 96–8). First of all, there is no reason that support for a conjunction can be factored into support for the conjuncts. Second, there are many different ways to express H as a conjunction, one conjunct of which is deductively entailed by E, and for some of those there clearly can be positive support (say, if H itself is the other conjunct). And perhaps most embarrassingly, Popper seems to have made the same mistake he pointed out for Carnap three decades earlier – by trying to factor the support into two components, he gets an understanding of $P(H \mid E)$, but doesn't show whether it is an increase or decrease from $P(H)$. So this apparent problem is not a serious one for Bayesian confirmation theory, though it does point out the fact that inductive and deductive support are hard to tell apart. The observation of a given raven as being black gives some minor deductive support to the hypothesis that all ravens are black (by showing that one potential deductive falsifier doesn't falsify it), but the more interesting question is what inductive support it gives to the claim that *other* ravens are black.

2.2. LOGICAL OMNISCIENCE

A far more serious worry for Bayesianism is the problem of logical omniscience. The problem is just that rational agents are often uncertain about logical truths – consider logic students who don't recognize a claim as a tautology, or mathematicians wondering about whether a given conjecture follows from the axioms. However, the standard probability axioms seem to entail that any logical truth must get probability 1, and thus be certain. I say 'seem to entail' because Kolmogorov's axioms stated in the abstract way that I gave above do not actually entail this claim – however, if we take the normal interpretation, and state that the set of possibilities is some set of (logically or metaphysically) possible worlds, and identify propositions with the set of worlds in which they are true, then logical (or even metaphysical!) omniscience does follow. We might try to reject this standard interpretation of the set of worlds, but this then raises problems for the connection between the set-theoretic operations of complementation and intersection (which the axioms deal with) and the semantic operations of negation and conjunction (which are needed in the interpretation). An early approach of this sort is that of (Hacking 1967), which uses a notion of 'personal possibility'.

Whatever the problems for the Kolmogorov axioms, the problem for more syntactic axiomatizations of probability is worse, since these versions clearly do entail logical omniscience, and the prospects for changing the interpretation of the theory are less clear. (Garber 1983) does give (as part of an attempt to resolve the old evidence problem) a way to change the interpretation that has some potential. He suggests that although the agent uses one language, there is no reason for the Bayesian theorist to represent the agent's credences by a probability function defined on that very language. Instead, we can treat each sentence of the agent's language as a sentential atom and apply the probability axioms to the language for sentential logic built on those atoms. This makes every sentence of the agent's language (including logical truths) come out contingent, so none of them are required to have probability 1. However, it preserves logical omniscience in the new language, and it raises the question of why the theorist's language omits the logical structure of the agent's language, even where the agent recognizes it.

Another, more sophisticated approach comes from (Gaifman 2004), who preserves the agent's language but gives a different syntactic axiomatization of probability theory, based on a designated finite fragment of the agent's language. He shows that under his axioms, the agent is only required to be logically omniscient about statements that have a deductive proof, every one of whose steps is in the designated fragment. If the fragment is taken to be the sentences short enough for the agent to grasp, then there will generally be many sentences that are themselves short enough to grasp, but all of whose proofs require steps that are too long to grasp. Thus, full logical omniscience is avoided. However, there seem to be cases where the problem is that the agent has never considered each step in a proof in the right order, and not that the agent can't grasp the individual steps. Gaifman's approach does nothing to deal with this sort of non-omniscience.

Yet another possible alternate approach is to say that Bayesian theory is an attempt to model ideal agents, but that actual scientists and mathematicians are non-ideal, and so there is no problem. This response is problematic to the extent that the lack of logical omniscience figures in the application of Bayesian confirmation theory to understanding the practices of actual (as opposed to ideal) scientists, as in some responses to the problem of old evidence (see below). But if these applications are less important, then this is not a significant problem. See chapter 6 of (Christensen 2004) for more discussion of this response.

Thus, there are several approaches to this problem, but there is as yet no fully satisfactory one. I suspect that approaches based on modifying the interpretation of the set in Kolmogorov's axiomatization will be more promising than more syntactic approaches, but this remains to be seen.

2.3. Old evidence

Another problem for Bayesianism, originally discussed in (Glymour 1980), is the problem of old evidence. The problem as it was originally posed is that by the standard Bayesian axioms, if $P(E) = 1$, then $P(H \mid E) = P(H)$, and so E can't confirm H. However, in actual scientific practice it seems that old evidence can confirm new hypotheses. The standard example is one of the historically most important pieces of support for Einstein's theory of general relativity. The perihelion shift of Mercury had long been known to disagree with the value predicted by Newtonian theory, and Einstein eventually noticed that it agreed with the value predicted by his theory. But since Einstein had long known the value of the perihelion shift of Mercury, he must already have assigned it probability 1, and thus the standard Bayesian account says that it could not have confirmed his hypothesis.

One approach to this problem, endorsed by (Garber 1983), is to concede that the perihelion shift itself gave no confirmation for Einstein's theory, but to claim that *the fact that Einstein's theory entailed the perihelion shift* gave the confirmation. Thus, the confirmation comes from a piece of logical learning, rather than from the old evidence. This approach clearly requires a solution to the problem of logical omniscience, but since this problem needs solution anyway, it reduces two problems for Bayesianism to one.

However, as pointed out by (Eells 1985), this example is only one type of old evidence. For Einstein, we could make a reasonable case that the new logical information was in fact the piece of evidence that confirmed his theory, rather than the old perihelion shift data. But for modern scientists, the logical fact and the perihelion shift *both* have probability 1, so that neither can do the confirmation. Eells calls this the problem of 'old new evidence' (that is, evidence that's already known, but was recognized as relevant when it was new) while Einstein's case was the problem of 'new old evidence' (evidence that's already known, but is only newly recognized as relevant to a theory). This problem of old new evidence (and the related problem of old old evidence, which is just a historical case of 'new old evidence') points out that Bayesian confirmation theory doesn't seem to be able to account for the confirmational relations among evidence and hypotheses that scientists now take to be well-established.

A first response to this problem is to consider the time at which the putative evidence was new, and to see whether it confirmed the hypothesis with respect to the historical probability function. However, there's no reason to suppose that these confirmation relations have remained the same – a piece of evidence that was seen as monumental at the time may now be seen as irrelevant in light of later developments, or vice versa. So an approach is needed that somehow appeals to the present, rather than to the time when the evidence was new.

One approach is to return to the question of which probability function P should be used in the inequality $P(H \mid E) > P(H)$. Rather than using the agent's actual degree of belief function (which has the problem that $P(E) = 1$), one might use the degree of belief function the agent *would* have had, if she hadn't already known E. However, as pointed out in (Earman 1992, 123), there may be no unique such function, and the function may differ from the actual one in confounding ways.

A more extreme version of this position is endorsed in (Williamson 2002) as well as (Carnap 1950), where they suggest that P is in fact some sort of hypothetical prior the agent would have had in the absence of any empirical knowledge whatsoever. However, many Bayesians find these functions even more problematic than the hypothetical ones removing just a single piece of information, and this suggestion seems to ignore the fact that much evidence is only evidence in light of our other background knowledge.

Another approach taken by some Bayesians is to endorse only Jeffrey conditionalization and not strict conditionalization – thus, even though the evidence may be old in some sense, it won't have probability 1, and thus we can have $P(H \mid E) \neq P(H)$. There is still a question of which measure of confirmation to use, as several measures guarantee that when $P(E)$ is close to 1 (as seems natural for old evidence even on this picture) there is almost no confirmation. However, (Christensen 1999) points out that other measures can do the right job here, so that if we take a pluralist picture, then this approach seems to work.[4] But this still seems to make the evidential relations *already* holding between an agent's beliefs dependent on how *future* increases in $P(E)$ would affect $P(H)$, which seems to get something conceptually wrong.

Thus, there is still work to be done on the problem of old evidence, though a variety of approaches allow most Bayesians not to worry about it.

2.4. NEW THEORIES

These previous two problems bring out the point that Bayesian confirmation theory seems to have no way to account for the effect of the development of a new theory. In a sense, the logical omniscience assumed in Bayesian theory goes beyond assigning probability 1 to tautologies, and also requires assigning probabilities to every sentence whatsoever, regardless of which ones the agent may have actually ever considered. The problem of old evidence is a serious problem because there seems to be a difference between the evidential relation between an old piece of evidence that a new hypothesis happens to explain (like in Einstein's case), and an old piece of evidence that a new hypothesis was specifically designed to accommodate – somehow the logical entailment relation between the hypothesis and evidence seems like it should be much more relevant in one case than the other, even though the theory had no prior before it was thought up. It also often seems that the introduction of a new theory induces a change in degrees of belief that can't be achieved by conditionalization or Jeffrey conditionalization, so that a new form of Bayesian update must be proposed. Thus, if Bayesianism is intended to help us understand actual scientific practice, rather than an extremely idealized version of it, a lot must be done to explain the role the introduction of new theories plays in science. (For an attempt to resolve this problem, see (Maher 1995)).

2.5. PROBLEM OF THE PRIORS

This brings us to perhaps the biggest problem facing Bayesian confirmation theory, which is the problem of the priors. Bayes' theorem shows that $P(H \mid E) = \frac{P(E \mid H)}{P(E)} P(H)$, so that whether $P(H \mid E) > P(H)$ just depends on whether $P(E \mid H) > P(E)$. Although some argument may be made for an objective assignment of value for $P(E \mid H)$ (for instance, many scientific theories explicitly attach probabilities to various observational outcomes, and it seems reasonable for a scientist who knows these values to take them as her likelihood), there really doesn't seem to be an objective way to specify the prior $P(E)$.

In fact, most Bayesians argue that many different priors are all permissible, whether the agents have the same background evidence or not. Thus, the picture given by Bayesian confirmation theory seems to indicate that the same evidence can provide confirmation for a hypothesis for one scientist, but disconfirmation of the same hypothesis for a different scientist. If this is right, then it seems to undermine the objectivity of science.

The responses to this problem lead to some of the most fundamental divisions among Bayesians. Some Bayesians are quite impressed by this problem and thus suggest that there should in fact be a unique objective set of prior probabilities that it is rational to have in any situation. Such 'objective Bayesians' (who seem to dominate among Bayesians in physics and the other sciences) often prefer the Cox argument over the other justifications of Bayesianism, because it presupposes their contention that there are unique priors to be had, and can then give values for this prior in many cases. In particular, it supports the Principle of Indifference, saying that rational agents should divide their credences equally among possibilities if they have no evidence favoring one over another.

However, 'subjective Bayesians', who tend to dominate among Bayesians in philosophy, point out that the Principle of Indifference had already been shown in (Bertrand 1889) to have serious problems, even before the birth of modern Bayesianism. The problem known as 'Bertrand's paradox' takes many different forms, but the central issue is the same – when there are infinitely many possibilities, there are different incompatible ways to divide them up where applying the Principle of Indifference gives different results. Paraphrasing a statement of the paradox from (van Fraassen 1989), we can imagine knowing only that a particular cubical box has side length somewhere between 1 and 3 meters long. Then it seems that we should assign credence 1/2 to its side length being between 1 and 2 meters long. But we could equally well describe the situation in terms of the surface area of a side being between 1 and 9 square meters, in which case it seems we should assign credence 3/8 to it being between 1 and 4 square meters. But then we have assigned two different credences to equivalent propositions, which is absurd. Jaynes tries to give a resolution to the specific example Bertrand gives with chords in a circle in (Jaynes 1973), but even he states that objective Bayesians must still do a lot of work to find the correct priors for other situations. (There is further interesting discussion of this paradox in (Joyce 2005, 167–71)).

Subjective Bayesians thus deny that we should adopt a general Principal of Indifference (though most admit that it is often quite useful in particular cases, when applied carefully), and so suggest that there are generally many different priors that are permissible for rational agents to have in a given situation. Thus, they must say something about the fact that scientists generally tend to agree as to which pieces of evidence confirm which hypotheses, since different priors will often give different confirmation relations.

A set of results that Bayesians often appeal to shows that for any two agents with different priors, as long as neither one assigns credence 0 to anything that the other doesn't and they update by sequentially conditionalizing on the same pieces of evidence, then their credence functions will eventually become arbitrarily close in the limit. (See Chapter 6 of (Earman 1992) and section 4 of (Hawthorne 2008) for more details on these theorems.) Some care needs to be taken in saying what measure of closeness will be used and what sequences of evidence will be considered, but the results are often taken to suggest that since most scientists have been exposed to large amounts of the same information (as part of their scientific education, and from reading published results of the same famous experiments), the credences of scientists will be close enough that they will tend to agree on the confirmation relations. The fact that there are some controversies can in fact be seen to support this position over objective Bayesianism. However, nothing in the results shows how quickly convergence of priors will happen unless some bound can be given on the original spread of priors, so the results can only be used in a very suggestive way. Thus, the problem of the priors still remains as perhaps the most significant challenge for Bayesian confirmation theory.

3. Bayesian Epistemology

In addition to its prominent role in confirmation theory, Bayesianism has become a prominent position in epistemology as well. However, since Bayesian epistemology focuses on the graded notions of degree of belief and confirmation, while traditional epistemology focuses on the all-or-nothing notions of belief, justification, and knowledge, some work must be done to say how these two sets of ideas relate.

One response endorsed by a few Bayesians is just to deny that the traditional concepts are useful at all – they are just informal notions that should be replaced for technical purposes by a formal notion, just as 'hot', 'cold', and 'warm' are just approximations that should be replaced by the formal notion of temperature. However, there are many arguments (such as those in (Williamson 2002)) that the concept of knowledge at least plays an important explanatory role, and that it can't just be disposed of. Also, most Bayesians seem to make use of some sort of notion of belief in explaining what it is to conditionalize on a proposition. However, some will follow (de Finetti 1931), insisting that a notion of 'belief', as opposed to 'degree of belief', must correspond to probability 1, and thus be unrevisable. Instead, one should have non-extreme degrees of belief for everything, and do everything in purely probabilistic terms. This position was endorsed by Richard Jeffrey as 'radical probabilism', and he replaced conditionalization with Jeffrey updating to avoid the update problem. However, it is exceedingly hard to avoid using any non-Bayesian notion in one's epistemology.

If we follow this idea, then many traditional problems of epistemology can be either solved or dissolved. Although the problem of Cartesian skepticism (roughly, that we could be radically deceived about the entire world, apart perhaps from a few very basic truths like 'I think, therefore I am') seems to suggest that we can never be certain of anything, this is quite compatible in radical probabilism with the claim that we can rationally have very high degrees of belief in many statements.

Hume's problem of induction seems to have an even better apparent resolution for the Bayesian. (This is a simplification of some of de Finetti's work on the notion of 'exchangeability', which is a way to make sense of the notion of independent identically distributed variables, even if one thinks the notion of an objective distribution for the variables makes no sense.) Consider some universal hypothesis H (say, that all ravens are black) to which an agent gives a non-zero prior, and consider a sequence of statements $E_1,E_2,E_3,\ldots$ that are all logically entailed by H. Then upon learning $E_1,\ldots,E_n$, conditionalization and Bayes' theorem show that one's new credence in H should be given by $P(H)/P(E_1\&\ldots\&E_n)$. But since $P(E_1\&\ldots\&E_n) = P(E_1)P(E_2 \mid E_1)\ldots P(E_n \mid E_1\&\ldots\&E_{n-1})$, we see that the posterior of H will eventually go above 1 (which of course is impossible) unless the $P(E_n \mid E_1\&\ldots\&E_{n-1})$ converge upwards to 1. Thus, the Bayesian can show that whatever one's credences, as long as one has a non-zero prior in a universal hypothesis, instances of this hypothesis will eventually give each other arbitrarily strong support. This proof gives no indication of the speed with which the sequence must converge to 1, but this seems exactly right. The requirement that H have a non-zero prior also turns out not to be innocuous, because it is just this assumption that prevents a problem with Goodman's 'new riddle of induction'. It turns out that if n observations of green emeralds are sufficient to give the agent credence 0.99 that the next emerald will be green, then she must have initially assigned credence at most 0.01 to any hypothesis of the form 'all emeralds are grue$_m$' for $m > n$, where 'grue$_m$' means 'green and one of the first m, or blue and one of the later emeralds'. Induction is saved without getting an outright contradiction from the new riddle, but the question of how to actually deal with grue is pushed into the priors.

Although it may be tempting for some Bayesians to entirely dismiss the traditional notions of belief, justification, and knowledge, most Bayesians admit that there should instead be some interaction between them. In particular, the notions of full belief and degree of belief seem like they should bear some particularly close connection. One popular idea, called the 'Lockean thesis' in (Foley 1993), states that full belief just is having a degree of belief above some threshold t.

As tempting as this picture might be, it has some serious problems. One particular type of case is the lottery paradox of Henry Kyburg. If there were some threshold $t < 1$ such that having degree of belief at least t suffices for full belief in a proposition, then we can consider a fair lottery with more than $1/(1 - t)$ tickets, of which the agent has exactly one. She ought to believe to degree greater than t that this ticket is not a winner, and thus the Lockean thesis says that she ought to believe her ticket will not win. However, many people share the intuition that in these sorts of cases, she doesn't *know* that the ticket won't win, and therefore she ought not to believe that.[5] The only way to avoid this while maintaining the Lockean thesis is to set the threshold for full belief equal to 1 – but then it seems that we can't count as believing very many of the things we ordinarily believe, since we normally allow for some slight positive chance of falsehood.

Thus, it seems that the connection between degree of belief and full belief must be somewhat subtle. (Kyburg 1988) suggests that the threshold for full belief may depend on the stakes that are relevant for decisions relating to the proposition in question. (Hawthorne and Weatherson 2000) argues that the threshold should be 1, but that we need to understand agents in terms of contextually specified probability functions, rather than a unified one, so that probability 1 will often occur in ordinary situations. Other proposals can and should be considered.

Similar issues arise in the connection between justification and confirmation. It seems very plausible to suggest that in order for a belief in H to be justified by a belief in E, H must be confirmed by E. But (Weatherson 2007) suggests that some standard anti-skeptical positions in epistemology that he calls 'dogmatist' must reject this principle. There have also been proposals like that in (Roush 2005) to link Bayesian notions with the notion of knowledge in order to deal with some of the traditional problems of epistemology.

4. Bayesian Statistics

Since Bayesianism claims to be the proper mathematical refinement of epistemology, and statistics is a sort of mathematical applied epistemology for the sciences, it's no surprise that Bayesianism has been extremely influential in statistics. However, to say that this influence has been controversial is an understatement – perhaps the biggest debate in statistics (and also in many disciplines where statistics is applied) is whether Bayesian statistics is better justified than the traditional 'frequentist' or 'classical' statistics.

The debate at its root comes down to one about the interpretation of probability, even though the dispute is not often argued at this level. The frequentist is not interested in anyone's subjective degrees of belief in a proposition – she says the goal of statistical methodology is to provide a strategy for deciding whether to accept or reject a hypothesis that is very likely to get things right. The way she makes sense of the 'very likely' in this claim is to consider repeating the test many times and see what fraction of those times we would make the wrong decision. Since the hypothesis either is true or is false (the frequentist doesn't care about how strongly anyone believes it's true or false, and the hypothesis is just a single case, so there's no meaningful frequency of truth or falsehood) this means that our test should reject a true hypothesis no more than (say) 5% of the time, and also accept a false hypothesis no more than 5% of the time. In many cases, it would take more data than we could gather to get a test that is this good, so they often make the choice between rejecting the hypothesis, and failing to reject the hypothesis, rather than accepting it. For this sort of test, they tend to try to make sure that we reject a true hypothesis in no more than 5% of cases, and then make the probability of rejecting a false hypothesis (called the 'power' of the test) as high as possible.

As an example, a farmer may be interested in whether a new farming method will make her plants grow larger than her old method. She can apply the new method to 100 plants, and the old method to 100 plants, and then measure their sizes. By looking at the standard deviation of all the sizes, and the difference in mean sizes for the two samples, she can calculate the chance that the means of two samples would be at least this different if there was only a single underlying distribution. (This chance is often called the 'p-value' of the samples.) If this p-value is no more than 5% (called the 'significance level' of the test as a whole), then she rejects the hypothesis of no difference, and accepts the hypothesis that the farming method made a difference. (The frequentist will tell her not to make any inference of the *amount* of difference the method made, because any such inference will increase her chances of being wrong beyond 5%. Strictly speaking, she shouldn't conclude anything from the p-value either – a smaller p-value *would* have been important if she had been performing a different test, but all she should pay attention to is what the highest *p*-value is that would have led her to reject the null hypothesis, because *that* is the probability that she would have made a mistake.)

The Bayesian by contrast says that this distinction between rejecting and failing to reject a hypothesis is a very artificial one – although these methods are unlikely to lead us to false conclusions, they don't tell us anything about how confident we should be in our conclusions. Bayesian methodology tells us exactly how confident we should be in any proposition given our evidence and our prior. Since no hypothesis is ever rejected or accepted, there is no need to worry about probability of error. Additionally, the methodology lets us pay attention to the full evidence, and not worry about what counterfactual evidence we *could* have gotten that would have led us to reject the hypothesis. So a farmer whose *p*-value is substantially lower than the significance level can be more confident in the hypothesis than one whose *p*-value just barely meets the threshold, and a farmer whose *p*-value almost meets the threshold can legitimately gain some confidence of difference. Furthermore, even the frequentist must admit that in many cases the exact truth of the null hypothesis is extremely unlikely (how plausible is it that two methods of growing plants are *exactly* equally good?) – thus, the frequentist is committed to distinguishing between 'rejecting' the hypothesis and disbelieving it, since a reasonable person should disbelieve it even in the absence of evidence sufficient for a frequentist to reject it.

In practice however, the different methodologies aren't so cleanly separated. Frequentists often do gain confidence from small *p*-values, even though their interpretation of their procedure gives them no justification for this. Bayesians often do accept or reject hypotheses, even though their method gives no bound on the chance of making mistakes. It seems that resolving the debates will require attention to the relationship between the attitudes of degree of belief and acceptance or rejection of hypotheses, as well as sensitivity to the actual goals that scientists want to achieve by means of statistical testing. More on this can be found in volumes like (Howson and Urbach 1989) and (Royall 1997).

5. Alternatives to Bayesianism

One thing that tends to make Bayesianism in different disciplines feel like a very different position is the fact that the anti-Bayesians in different fields are very different. Thus, Bayesians in one discipline might endorse modifications of the theory that sound completely anti-Bayesian to workers in another discipline. For confirmation theory, the primary contenders are either older and decisively refuted theories like Hempel's hypothetico-deductivism (Hempel 1945) or modifications of Bayesianism that raise as many problems as they solve. (A few examples are possibility theory (Zadeh 1978), Dempster-Shafer theory (Shafer 1976) and 'mushy credences' (White 2009). Many of these alternatives are discussed in (Walley 1991) and (Halpern 2003)). In epistemology, the main alternative seems to be to avoid dealing with the comparative notions of degree of belief altogether, focusing instead on just the traditional notions of full belief, justification, and knowledge. In statistics, the traditional frequentist methodology seems to have significant foundational problems worse than the Bayesian problem of the priors – perhaps some third alternative, or fusion of the methodologies is needed, but the dispute is clearly an important one.

Thus, Bayesianism is a position that lies at the nexus of many different disputes in different fields. Whether or not Bayesianism (or some modification of it) is the correct approach to any of these sets of problems, it allows us to bring to bear techniques developed in these other areas, and has been a very fruitful research program.

Short Biography

Kenny Easwaran received his PhD from the Group in Logic and the Methodology of Science at UC Berkeley in May 2008, with a dissertation on the appropriate mathematical formalism for Bayesian conditional probability. He is currently assistant professor at the University of Southern California. He also works on the philosophy of mathematics in addition to formal epistemology.

Notes

* Correspondence: USC School of Philosophy, University of Southern California, Mudd Hall of Philosophy, 3709 Trousdale Pkwy, Los Angeles, CA 90089, USA. Email: easwaran@gmail.com.

[1] Bayes' theorem is often seen as so central to Bayesian confirmation theory that people assume that the entire enterprise is named after the theorem – but (Earman 1992, ch. 1) argues that the other aspects of the paper in which the theorem was found really do justify naming the enterprise after the author of that paper.

[2] There is one slight qualification – if either $P(H)$ or $P(E)$ is 0 or 1, then some of these measures may be undefined, or may depend on the specific account of conditional probability involved. In these cases, which will be relevant in Section 2.3, there can be disagreement among the measures.

[3] The measures based on ratios instead of differences indicate evidential neutrality with a value of 1 instead of 0 – some authors take the logarithm of these measures, to return to a traditional additive scale centered on 0.

[4] Joyce (1999) suggests that this kind of pluralism will allow for a solution of the old evidence problem even if $P(E) = 1$, if we use a notion of conditional probability that can define $P(H \mid \neg E)$ even when $P(\neg E) = 0$. (Such notions date back at least to (Popper 1959), and are discussed extensively in (Easwaran 2008)). However, there is no convincing reason for $P(\neg E)$ to equal 0 unless it is taken as an epistemic impossibility by the agent, and every account of probability conditional on events of probability 0 would make $P(H \mid \neg E) = 1$ when $\neg E$ is an epistemic impossibility. Thus, this approach would mean that old evidence could only disconfirm, and never confirm a hypothesis, which seems not to be a solution at all.

[5] We might instead point to the fact that her belief should apply to every ticket, which leads to an inconsistency – but precisely this possibility is often seen as a *strength* of Bayesianism in responding to the preface paradox.

Works Cited

Bertrand, J. *Calcul des probabilités*. Paris: Gauthier-Villars, 1889.

Carnap, R. 'Two concepts of probability.' *Philosophy and Phenomenological Research* 5.4 (1945): 513–32.

——. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 1950.

Christensen, D. 'Measuring confirmation.' *The Journal of Philosophy* 96.9 (1999): 437–61.

——. *Putting Logic in its Place: Formal Constraints on Rational Belief*. Oxford: Oxford University Press, 2004.

Earman, J. *Bayes or Bust?* Cambridge, MA: The MIT Press, 1992.

Easwaran, K. *The Foundations of Conditional Probability*. PhD thesis, Berkeley: University of California, 2008.

Eells, E. 'Problems of old evidence.' *Pacific Philosophical Quarterly* 66 (1985): 283–302.

de Finetti, B. 'Probabilismo.' *Logos* 14 (1931): 163–219.

Fitelson, B. 'The plurality of Bayesian measures of confirmation and the problem of measure sensitivity.' *Philosophy of Science* 66.3 (1999): S362–78.

——. 'The paradox of confirmation.' *Philosophy Compass* 1.1 (2006): 95–113.

Foley, R. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford: Oxford University Press, 1993.

van Fraassen, B. *Laws and Symmetry*. Oxford: Clarendon Press, 1989.

Gaifman, H. 'Reasoning with limited resources and assigning probabilities to arithmetical statements.' *Synthese* 140 (2004): 97–119.

Garber, D. 'Old evidence and logical omniscience in Bayesian confirmation theory.' *Testing Scientific Theories*. Ed. J. Earman. Minneapolis: Minnesota Studies in the Philosophy of Science, 1983. 99–132.

Glymour, C. *Theory and Evidence*. Princeton: Princeton University Press, 1980.

Hacking, I. 'Slightly more realistic personal probability.' *Philosophy of Science* 34.4 (1967): 311–25.

Halpern, J. *Reasoning About Uncertainty*. Cambridge, MA: MIT Press, 2003.

Hawthorne, J. 'Inductive logic.' *Stanford Encyclopedia of Philosophy* (2008). 19 Feb. 2011 <http://plato.stanford.edu/entries/logic-inductive/>.

—— and B. Weatherson, 'Beliefs old and new.' 19 Feb. 2011 unpublished—Available at: <http://brian.weatherson.org/belief.pdf>, 2000.

Hempel, C. 'Studies in the logic of confirmation (i.).' *Mind* 54.213 (1945): 1–26.
Howson, C. and P. Urbach. *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court, 1989.
Jaynes, E. T. 'The well-posed problem.' *Foundations of Physics* 3.4 (1973): 477–92.
Joyce, J. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press, 1999.
——. 'How probabilities reflect evidence.' *Philosophical Perspectives* 19 (2005): 153–78.
Kyburg, H. 'Full belief.' *Theory and Decision* 25 (1988): 137–62.
Maher, P. 'Probabilities for new theories.' *Philosophical Studies* 77 (1995): 103–15.
Popper, K. 'Degree of confirmation.' *The British Journal for the Philosophy of Science* 5 (1954): 143–9.
——. *The Logic of Scientific Discovery*. chapter iv*, pp. 326–48. New York: Harper & Row, 1959.
—— and D. Miller. 'A proof of the impossibility of inductive probability.' *Nature* 302 (1983): 687–8.
Roush, S. *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Oxford University Press, 2005.
Royall, R. *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall, 1997.
Shafer, G. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.
Walley, P. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall, 1991.
Weatherson, B. 'The Bayesian and the dogmatist.' *Proceedings of the Aristotelian Society* 107 (2007): 169–85.
White, R. 'Evidential symmetry and mushy credence.' *Oxford Studies in Epistemology* 3 (2009): 161–88.
Williamson, T. *Knowledge and its Limits*. Oxford: Oxford University Press, 2002.
Zadeh, L. 'Fuzzy sets as the basis for a theory of possibility.' *Fuzzy Sets and Systems* 1.1 (1978): 3–28.