Theoretical Omniscience: Old Evidence or New Theory

André C. R. Martins

Data Analysis Group
Escola de Artes Ciências e Humanidades
Universidade de São Paulo
Av. Arlindo Bétio, 1000
Parque Ecológico do Tietê – São Paulo
CEP: 03828-080 – SP
Brazil
amartins@usp.br
Phone: (55)(11) 6545.0243

Abstract

I will show that, in the Problem of Old Evidence, unless a rational agent has a property I will call theoretical omniscience (a stronger version of logical omniscience), a problem with non-commutativity of the learning theories follows. Therefore, scientists, when trying to behave as close to rationality as possible, should behave in a way close to the counterfactual strategy. The concept of theoretical omniscience will be applied to the problem of Jeffrey conditionalization, as an example, and we will see that a more complete theoretical model can provide a classical conditionalization where you can learn that data was wrong and all you will not unlearn is your memory.

Keywords: Bayesian Confirmation Theory, Theoretical Omniscience, Rationality, Jeffrey Conditionalization

1. Introduction

Bayesianism tells us that scientists update their beliefs about different theories that describe the world by assigning probabilities to each possibility and by updating those probabilities, when faced with new data, by using the rules of Bayesian Statistics. Having normative quantitative methods for induction is certainly a good idea and can provide predictions and testable models of the way Science works. The question becomes if Bayesian methods is what one should really do. To answer that, we need to know which characteristics are desirable in an inductive logic and check what those characteristics dictate to any measure of plausibility of a proposition one wishes to adopt. It is reasonable to start with a small number of desired properties. First, the measure of plausibility should be a number. Second, its results should be equivalent to our common sense in the simple cases where we can trust it to tell us that a proposition has become more or less plausible. At last, the results should be coherent, that is, if there is more than one way to arrive at plausibilities for one proposition, all of them should provide the same result when the calculation is performed. Jaynes (2003) demonstrates that, if one requires that a measure of the plausibility should obey these three desiderata, one can arrive at some rules that those measures should obey. Different measures of plausibility are certainly possible, since the demonstration only fixes them up to a monotonic transformation and one of the possible measures obeys the theorems of probability and, in that case, the plausibilities must be altered according to Bayes' Theorem. Given that it is natural to think of the plausibilities as how probable one believes a theory to be the best one available, this choice of the measure is the natural one and we are normatively forced to use Bayesian methods if one wants to use an inductive logic.

However, although it is recognized that Bayesianism has several merits, there are still some perceived problems with its application to the problem of theory choice (see Kaplan, 2002 for a review of several objections). One of the problems was appointed by Glymour (1980) and is based on what happens when one is faced with a new theory that is confirmed not by evidence gathered after the theory is developed, but before. The classical example in the discussion of this problem is that of advance of the perihelion of Mercury. The argument goes like this: when Einstein proposed the General Relativity Theory as a description of the structure of space-time and, therefore, was able to predict the movements of planets using his new theory, the problem with the advance of the perihelion of Mercury was known for a long time. Newtonian Mechanics was not capable of explaining that behavior, but the behavior was already known and confirmed by observation. Therefore, the probability associated with the advance should be taken to be 1 (or, at least, very close to 1, since there is always the possibility of some yet undetected mistake or misinterpretation of the data). However, if one tries to update the probability associated to General Relativity by using an observation that one is certain to obtain, the prior probability would not be changed and, therefore, there would be no confirmation to the theory coming from that observation, despite the fact that it was the only theory available capable of explaining the observations. And that is certainly wrong, since the advance of the perihelion should confirm General Relativity, as it actually did. At first sight, it seems as if there was something wrong with Bayesianism, if no old evidence can be ever used to confirm a new theory. This problem is known as the Problem of Old Evidence, POE. Should this analysis be true, it would pose a serious problem to any proponents of Bayesianism and, therefore, a solution to this problem was needed.

Several people have suggested ways to deal with this apparent problem. Garber (1983) suggested that we should relax the assumption of logical omniscience, by pointing that, when we have a new theory, we should condition the probability of the theory not on the result of the experiment, since doing that would mean learning nothing, but on the new knowledge that the theory implies the experimental result. Earman (1992, see also Fitelson, 2004), on the other hand, has proposed, and abandoned the idea, that one can work around this problem if one does not assign a probability of 1 to the old evidence, since, as mentioned above, there is always the possibility of some kind of error. If, when updating your probabilities, you use an observation that you are only 0.999 sure about, you can still learn something and change the values of the probabilities you associate with the theory that predicts it. You will not learn as much as you could from a surprising event (one with a much lower probability), in the sense that your probability will not change as much as it would for the surprising event, but you will still learn. Another attempt to solve the problem is the counterfactual strategy, proposed by Howson and Urbach (1993). The idea is that one should exclude the knowledge of the evidence E and predict it, by using the new theory and, only after that, update the probabilities supposing that E has just been observed. The main criticism to this proposal (Glymour, 1980) is that forgetting that one knows E is something people might have difficult in doing and it is not clear how much knowledge you should actually ignore when partially forgetting your knowledge in order to apply the Bayesian methods.

In this article, I will present a different solution that basically leads to the counterfactual strategy but that, as we will see, has also some points in common with Earman approach. I will argue that the real problem is that, besides logical omniscience, Bayesianism requires that rational agents should also be theoretically omniscient. That is, a rational agent should have complete theoretical knowledge in the sense that every conceivable theory should be known to her before any data is ever collected. Given the impossibility of this, I will argue that the counterfactual strategy is the best available alternative and, despite the mentioned problems, it can be understood as a useful and quite often correct heuristics.

As an example of the consequences of not using theoretically complete models, I will also show that complete theoretical knowledge can help explain the problems with non-commutativeness of Jeffrey conditionalization. We will see that classical conditionalization, for a complete model, can have the features that have lead Jeffrey to propose his conditionalization, that is, it is possible to unlearn data, as long as we use a more complete model. As an example, I will analyze the problem of deciding whether every Scot wear kilts or not, based on an unreliable observation that one Scot might be wearing a kilt right now, as presented in Huber (2005). We will see that, if one describes the model better, considering every logical possibility, the probability in the full model associated with the reliability of the information behaves in the intuitively expected way, even though in the calculations done by Huber, a version of the problem of old evidence was observed.

2. Old Evidence and Theoretical Omniscience

One of the desiderata used by Jaynes to obtain the measure of plausibility was that, if there was more than one way to calculate the plausibility, both should provide the same result. The order one learns something should not matter to the final opinion of a rational agent, and it is a known fact that classical conditionalization is commutative. If you start with the same prior and learn the same facts, you will arrive at the same posterior, regardless of order. This is not true for Jeffrey's conditionalization and that has been a cause of concern about its use. We will return to the problem of Jeffrey's conditionalization later and, for now, I will deal with another problem, that of old evidence, where the order of the learning process, depending on how you solve the problem, can make a difference.

Suppose a certain observation E has been made in the past and, therefore, it is known that E. The problem is that E is not expected by any of the known theories (or, at least, was considered to be very unlikely). Later, a theory H is developed and it becomes clear that H predicts that E obtains. It is reasonable to think that the observation E confirms the theory H. That was exactly what was observed in the case of the advance of the perihelion of Mercury. The advance was a known fact by the time Einstein predicted it should occur by using the tools of the just created General Relativity. However, if one tries to condition H on an information E that already have a probability of P(E)=1 associated to it, there will be no change in the probability associated to H and, therefore, no confirmation whatsoever. The problem of old evidence is exactly to understand why and how E actually confirms H under these circumstances, as it is basically agreed that it should.

It is true that one should never assign a probability of exactly 1 to any proposition that is not a tautology, since any observation can later be shown to have been a mistake. Fitelson (2004) shows that, by taking a high value for P(E) but not exactly 1, if the likelihood is high enough, E will provide confirmation to H and, apparently, the problem becomes fixed, since one can have confirmation from old evidence. However, there is still a problem of non-commutativity associated with this. Take the two following scenarios, where the only difference is the order of the learning. The first scenario is just the one described above, where you learn about the new theory after you already know about the confirmatory evidence. In the second scenario, the order is reversed. You have learned about the theory and that the theory predicts E. Now, you have no idea if E will obtain and, since the experiment will still be performed, you have to choose a value for P(E) using your favorite method of assigning priors (I will not enter in the discussion of how to do it in this article). In this case, P(E) can be large or small, but we can say that it is certainly not 1 and therefore, when E is observed, it confirms theory H, as it should. It is still true that, the more surprising that E is, that is the smaller P(E) is, the higher the confirmation of H will be.

Notice that we have a problem. The update associated with E is given by the same expression in both scenarios.

$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)} \qquad (1)$$

Since the prior probability that H is true is also the same, as well as the probability of observing E if the theory H is correct, the only way that the posteriors will be the same in both scenarios is if one uses in both the same value for P(E). This is basically what the counterfactual strategy proposes and, requiring that the learning of the theory and the learning of the experiment should be commutative shows clearly how one should do the calculation.

Critics of the counterfactual strategy have claimed that it does not make it clear how much one should actually forget. The argument is that simply forgetting E does not necessarily means forgetting that you know how other scientists have reacted since the creation of the theory and knowing their opinions will affect your probabilities. But, if one is required to arrive at exactly the same conclusion, the procedure is simple. You should forget everything you have learned about the experiment. As a matter of fact, you have to redo all the calculations, assuming you have always known the theory, even before you have learned anything, directly or indirectly, about the experiment. The problem is that, in order to do that, it may seem that you should have perfect memory, so that you could remember all of your learning history associated to the problem, in order to recalculate and relearn everything.

Here, there are two different lines of reasoning we can follow, the normative and the descriptive. From the normative point of view, the result that one should either know all theories beforehand or have perfect memory in order to redo all his calculations foes follow. However, this is not feasible to any human or machine and, therefore, could not be an exact description of real behavior. Still, Bayesianism, when interpreted literally, should not be seen as a perfect description, but as a limiting case that, for most situations, should provide a close approximation. Notice that another of the criticisms of Bayesianism is that it requires people to be logically omniscient. That is, unless one is willing to face the possibility of arriving at results that are not coherent, one should always assign probability 1 to every tautology, no matter how complicated the reasoning behind it is, and probability 0 to every contradiction. But, of course, people do not always recognize propositions that are tautologies or contradictions as such and somebody might assign a different probability to those, representing the fact that one is not sure. That is, Bayesianism is known to be, at best, from the descriptive point of view, an approximation to the behavior of real scientists. At least, I would personally hope so, since I'd expect serious scientists to analyze their positions with great care and, therefore, that they should make up their minds with a dynamics not too far off that of an ideal rational agent. However, it is true that, depending on the circumstances, the match between Bayesianism and real behavior will not be perfect and it should not be expected to always be perfect.

Now, notice that, in order to solve the problem of old evidence, one does not necessarily need perfect memory. If, at first, you knew about the theory, there is no old evidence, since the theory is older and, therefore, there is no problem. That is, you only need to know every theory before the evidence supporting it ever appears. Or, in other words, rational people should have a version of logical omniscience that might be stronger than postulated before. Not only every logical truth should be known to them, but they should also know every possible and conceivable theory, no matter how unlikely that theory might be. It might seem that it is not clear that this version is really stronger, since by knowing every logical truth, it is possible to argue that every possible consequence of each proposition would be known and all theories would become accessible. Still, I believe this is a stronger version, especially if logical omniscience is understood as only the ability to see all consequences of a proposition in zero time. That ability would suffice for applications where no new theories are involved, but it does not include the ability to know every possible proposition one might ever come up with, something that the stronger version, which I call theoretical omniscience, does require.

Up to this point, there is nothing really new to it, since I have only proposed to trade one problem for another. Instead of old evidence, we have now a problem of the new theory, since no theory should ever be new. However, in my opinion, since we can at least drop the requirement of perfect memory, this is an improvement in old ideas. Instead of two impossibilities (logical omniscience and perfect memory), one is left with one (theoretical omniscience, that includes logical omniscience). It is still impossible, anyway. But it brings to light a characteristic of Bayesian problems that should be clear to anyone who tries to use Bayesian methods and, in that sense, I think it has nice pedagogical features when it shows that every effort made into a more complete theoretical description should be made before any problem is studied and new data learned. Therefore, in order to illustrate how this can at help, I will use the hint that we should always be looking for more complete theoretical models to solve a couple of puzzles in Bayesianism. Of course, since nobody is omniscient, what I will actually do is only to study the examples under a more complete model than previously done. We will see that, if one makes the analysis more complete, that is, closer to the case where all theories would be evaluated, the problems presented bellow can be better understood and solved.

Finally, it should be noticed that there will be cases where ignoring the existence of theories will lead to larger mistakes and cases where it is not so important, depending, obviously, on the likelihood of the data, given each possible theory and its prior. Fitelson and Thomason (2005) discusses in a different context of the one presented here, when approximating by a less complete theoretical background should provide no problem.

3. Jeffrey's Conditionalization or Memory

Classical conditionalization, where you calculate P(H|E) means that you are assuming that the evidence E that you are using has probability 1. The problem here is that giving probability 1 to a proposition is something that you should only do for tautologies or things you are so certain about you will never want to change your mind about. Experimental results, on the other hand, are propositions that might later be proven wrong. But, if you are strictly obeying the Bayes Theorem (as you should be, if you are a Bayesian), once you assign probability 1 or 0 to a proposition, you are stuck there and no amount of evidence would ever convince you otherwise. And, since any observation is never completely reliable, you should make your choices in a way that would allow you to later change your mind.

In order to solve this apparent paradox, Jeffrey (1983, 2004) proposed a different kind of conditionalization. Now, instead of assuming that E has probability 1, you should condition not on proposition E but on the change of your belief in the probability E. That is, the new probability of H, represented by *new*(H) to make it clear it is obtained by Jeffrey's conditionalization, would be given by

$$new(H) = p(H|E)\ p(E) + p(H|\neg E)\ p(\neg E) \qquad (2)$$

And, as long as P(E) never becomes exactly 0 or 1, you can always change it later and reapply the rule above to update your probabilities again. The problem is that if you make two consecutive conditionalizations, based on evidences E and F, using the same value in both cases for P(E) and P(F), but changing the order you learn each piece of evidence, you will arrive at different results. That is, Jeffrey's condionalization is not commutative. Since the order of learning is not any information about H at all, it shouldn't make any difference.

Let's try to understand what is going on by applying the idea that a Bayesian should, in principle, have theoretical omniscience when performing a classic conditionalization. Of course, I am certainly not theoretically omniscient, but I can propose, at least, a more complete theoretical description. Since it is certainly a possibility that the data we believe in might be wrong, we have to include that in the theoretical model. In other words, not only we need to know what H predicts, that is it implies E or whatever the value it assigns for p(E|H), but we also need a theoretical model for the observation of E. If one tells you that E was observed, or, if you make an experiment that seems to obtain E, what you do know for sure is B, that someone reported E or that you believe that E was obtained. B simply means that right now, it seems that E is true, a proposition much weaker than E and one you can safely assign certainty to. So you should actually condition on B, on your belief that E obtains, and, for that, you need a model that will predict you how likely you are to believe in E if E obtains, P(B|E) but also how likely you are to believe in E if we'd actually have ¬E as true, P(B|¬E).

Notice that the problem with Jeffrey conditionalization is that, when you have two propositions you want to reason about, A and B, when you condition them on a new value for A, say *new* (A) = 0.5, P(B) is also changed following Equation 2, but you actually have no control over that. If later you condition on *new* (B) = 0.7, P(A) will change to a different value and no longer 0.5, as it would be if the conditioning on A were the last thing you did. There is some interaction between both propositions that is assumed, but never made exactly clear. The problem is that saying that *new* (A) = 0.5 is not equivalent to a set of data. If initially, P(A) = 0.5, the data that would classically correspond to that change can be no data at all. If P(A) = 0.49, we can still have some weak data, but if, originally, P(A) = 0.1, the conditionalization is equivalent to a much stronger data. Since we have different intermediary values for the probabilities, each conditionalization is actually, from a classical and theoretically complete point of view, different, unless some extra conditions are imposed on how the conditionalization is performed.

Still, Jeffrey conditionalization can be understood as a special case of classical conditionalization, equivalent to some data that should be guessed from its effects. Per example, if, at first, you consider E to be unlikely to happen, say P(E) = 0.05, after learning B and supposing that the experiment was reliable enough so that you can assign the values P(B|E)=0.99 and P(B|¬E) = 0.01, you can easily obtain a new probability for E, by applying Bayes Theorem, that is P(E|B) = 0.839. That is, as in Jeffrey's conditionalization, we have no longer complete certainty about E and the knowledge of E can be unlearned. Jeffrey conditionalization would start by stating that P(E) was altered to 0.839 and condition on that value. If one is only interested about E, that is fine and should present no problem. However, if there are other propositions that we want to know, Jeffrey conditionalization will, without our control, introduce correlations between both propositions that might correspond or not to our real models. By describing the problem fully, closer to a theoretically omniscient description, the problems with not being able to unlearn E are gone and you can still control and learn about the relationship between the propositions.

What we won't be able to unlearn using classical conditionalization in the example above is B, that is that, at some point in time, we had reasons to believe E. But this certainty is not a problem and it is actually called memory. And now, since classical conditionalization is commutative, as long as you have a model for how unreliable each piece of information is, you don't have any problems with learning later that some data should be ignored. The only thing you won't be able to ignore is B, that you have once thought that E was true. So, a rational agent will be able to change his mind about anything regarding the real world, the only thing he won't be able to change his opinions about is about what he thought in the past.

Finally, it is interesting to notice that, when you introduce in the theory the possibility that E is not absolutely certain, some aspects of the proposal of Earman that the new theory have its associated probability updated by a large probability associated to E, instead of P(E)=1, are recovered, since we won't have certainty about the meaning of the data anymore. It is still not the same strategy, since we should have known that the theory H existed all along and the conditioning should be done following the counterfactual idea, since we are not theoretically omniscient.

4. Reliability of Information

Huber (2005) proposed that the less reliable a source of information is, the more confirmation one will get from it. Said that way, his affirmation sounds simply wrong to me and we will see that, although his mathematical result does hold, what he calls reliability should more properly be referred to as surprise. By changing his example to a more complete theoretical description of the process he studied, we will see that a measure of reliability naturally presents itself and it behaves in the expected way, besides illustrating better why more complete theories, closer to theoretical omniscience, matter.

The example in Huber dealt with checking the veracity of the theory H that all Scots wear kilts. In order to do that, the prior p(H) is changed to a new value by conditioning on uncertain information E that Stephen is wearing a kilt. By using Jeffrey conditionalization, Huber shows that the less reliable the information is, that is the smaller p(E) is, the larger the confirmation of H. In order to make this model closer to be theoretically complete, we have to introduce all possibilities we can think of. First, it is possible that H is true or not. If H is not true, Stephen might still be one of the Scots who wear kilts (W) or not (¬W). IF H is true, W is necessarily so. Again, if Scott wears kilts (either because every Scot does it or because he is one of those who do), there is a chance he is wearing it now (N) or not (¬N). If he does not wear it (¬W), then it is sure he is not wearing one now (¬N). Finally, since the observation is not reliable, if he is wearing now (N), the agent might make a visual observation that seems to show Stephen is wearing it (V) or she might see it wrong, thinking she sees he is not (¬V). The same is true in the case Stephen is not wearing kilts right now (¬N), but with different probabilities. It is reasonable to expect that, even if unreliable, the chance of believing Stephen is wearing kilts right now, V, should be larger if he is actually wearing than if not. This way, in order to make the equations a little easier to write, I define the following probabilities:

h = P(H), the prior probability H is true

s = P(W|¬H), if H is false, the chance Stephen still wears kilts.

n = P(N|WH) = P(N|W¬H), where I am assuming, to make it simpler, that the chance Stephen is wearing a kilt right now does not depend on H.

o = P(V|N), where the reliability is the same regardless of E or H.

q = P(V|¬N), another measure of reliability, associated with the other type of error.

Notice that, for those variables, what really counts as the reliability of the measure is both o and q. If o is large enough, if Stephen is wearing a kilt, the agent is very likely to think so. Also, if Stephen is not wearing a kilt, a small value of q means the agent will probably conclude correctly that he is not wearing it. Finally, since P(W|H)=1 and P(¬N|¬W)=1, those do not need to be included in the model. The real problem now becomes calculating P(H|V) and analyzing how it behaves if o increases and if q decreases.

By applying Bayes' Theorem to the above problem, we have

$$P(H|V) = \frac{hno+h(1-n)q}{(hn+sn-hsn)o+(hsn-hn-sn+1)q} \qquad (3)$$

Equation 3 has the general form

$$\frac{ao+bq}{co+dq}$$

Calculating the derivative with respect to o, it is easy to see that its sign depends on the sign of the term ad-bc. Therefore, by studying the sign of this term, one can easily determine whether P(H|V) increases or decreases with o. A similar result is valid for q, only the signs of the analysis change, since the meaning of a,b,c and d will be different and, therefore, I will omit the calculations for q here, only mentioning its result at the end. Returning to o, we can see that, after a few simplifications,

$$ad-bc = hn(1+hs-h-s)$$

That is, the sign of the partial derivative of o depends on the sign of the term in the parenthesis, g. It is easy to see that g(h,s) = 0 only when h = 1 or s = 1 and larger than 0 for every value of h and s that are under those limits. Therefore, the derivative never gets negative and is only 0 in the limits when one is sure that every Scot wears kilts (h = 1), or that Stephen wears kilts for sure, even if other Scots don't (s = 1). This is reasonable, since if you are already sure that Stephen wear kilts, making an observation where it seems he is wearing one now won't change your opinion much. Everywhere else, the partial derivative is positive, meaning that, if the reliability measured by o is larger, P(H|V) will also be. The term for q has the opposite sign, as it should be expected and therefore, the smaller the chance of making the mistake q measures, the larger P(H|V). Again, given a more complete theoretical model and the natural definitions of reliability in this model (there are two, associated with each possible type of error), the more reliable a result is, the more it confirms the hypothesis. There is nothing wrong with the calculations in Huber, but I have shown that P(E) shouldn't be called reliability, since it is not really associated with a model for error in observations. It can, however, be considered how surprising E is initially and, from there, Huber result does follow, except that the conclusion from his result becomes that surprising observations tell us more than expected ones, something that actually sounds very reasonable.

5. Discussion

We have seen that a rational agent should not only be logically omniscient, but also, theoretically omniscient, if she is to avoid commutative problems when one changes the order the data and the theory is learned. Since theoretical omniscience is not possible to achieve and we actually learn theories after some data, the behavior of a scientist who updates her believes as close as possible to normative Bayesianism should conform to the counterfactual strategy, that is, she must redo all the calculations about her opinions, supposing the theory was known from the beginning when she learns a new theory. This presents a different problem, in that it requires perfect memory in order to be able to redo the calculations from that point on, since the scientist will need to remember every piece of information that has changed her opinion since the data was learned. It seems reasonable to assume that nobody really does that, but probably adopt an approximation, where the confirmatory data is the only part of the knowledge that is actually considered unknown, until the probabilities are updated to include the new theory.

When applied to the problem of uncertain data, that is, data that we would like to condition on, but we are not absolutely certain about its truth content, the prescription that we should include complete models lead us to model that uncertainty in the data. The example presented above is certainly not complete and, therefore, far from being theoretically omniscient. It is easy to see how it could be made more complete, if necessary, per example, by allowing the reliabilities to have not a fixed value, but a prior distribution, so that we can also learn about the reliability as we perform the experiments. Still, even with that approximation, we were able to obtain a simple model where any data, other than the fact we once believed that we have made a specific observation, can later be unlearned. That was something that was not considered feasible for classical conditionalization, but we have seen it is just a matter of describing the process. I have shown that the traditional solution to this believed failure of classical conditionalization, Jeffrey conditionalization, is actually equivalent to classical conditionalization where the data is not completely trustworthy. But Jeffrey conditionalization is not equivalent to a fixed amount of data and hides relationships between the propositions that we have no control over and that is the reason of the commutative problems observed in the literature. That is, it can, at best, provide a good approximation to the way one should actually conditionalize. On the other hand, the main objection to classical conditionalization, that is, the fact that there are propositions you must condition on that will be assigned probability 1 to and, therefore, you will not be able to unlearn later, can be understood, when the more complete model is used, as the fact that a rational agent will not change his believes about his old epistemic states. That is, she will know that, once, she had obtained some evidence that seemed to indicate some data was actually true. And that is the only knowledge that she will not be able to change. Notice that, contrary to human beings, who might be irrational and convinced that something they believe was a delusion, a rational agent would not have this problem and, therefore, the stability of her memory is not, in my opinion, a problem at all.

References

Earman, J. (1992) *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MIT Press.

Fitelson, B. (2004) Earman on old evidence and measures of confirmation, manuscript

Fitelson, B. & Thomason, N. (2005) Bayesians sometimes cannot ignore even very implausible theories (even ones that have not yet been thought of), manuscript.

Garber, D. (1983) Old evidence and logical omniscience in Bayesian Confirmation Theory, *in* Earman, J. (ed.) *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, Vol. X*. Minneapolis, University of Minneapolis Press.

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

Howson, C. & Urbach, P. (1993) *Scientific Reasoning: The Bayesian Approach*. Chicago, Open Court.

Huber, F. (2005) Subjective probabilities as basis for scientific reasoning?. *Brit. J. Phil. Sci.* **56**, 101-116.

Jaynes, E.T. (2003) *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press).

Jeffrey, R. C. (1983) *The Logic of Decision*. Chicago, University of Chicago Press.

Jeffrey, R. C. (2004) *Subjective Probability: The Real Thing*. Cambridge, Cambridge University Press.

Kaplan, M. (2002) Decision Theory and Epistemology, *in* Moser, P. K. (ed.) *The Oxford Handbook of Epistemology*, Oxford, Oxford University Press.