Rerum Causae, 7 (2015), 82-104

The Problem of Old Evidence in Bayesian Confirmation Theory: A Solution Based on the Difference Between Update and Justification

Chloé de Canson*

Abstract

In this paper, I discuss the problem of old evidence in Bayesian Confirmation Theory and attempt to build on a suggestion by Lange (1999) to provide a satisfactory solution. I present the problem, argue that it cannot be solved by the standard appeal to counterfactuals, and show that a modification of the Bayesian framework along the lines of what Lange proposes is necessary. Then, within that new framework, I show how the problem of old evidence can be solved.

Introduction

The confirmation of scientific hypotheses is a central topic in philosophy of science. The main idea behind the different accounts of confirmation of scientific hypotheses is that pieces of evidence can lend support to theories, they

* For helpful discussions on the topic and comments on this paper, I am greatful to John Worrall, Olivier Roy, Luc Bovens and Alexandru Marcoci.

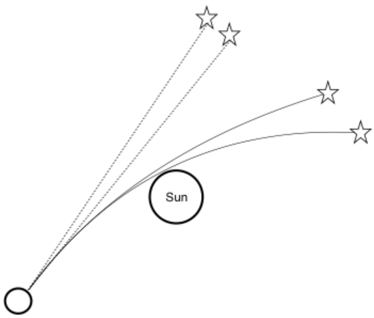

Figure 1: Eddington's Test

can confirm them. This is usually understood as meaning that, in light of some evidence, hypotheses can become more plausible. Let us illustrate this with an example. Let a hypothesis H be 'all ravens are black', and let the evidence E be the observation of a raven that is indeed black. Intuitively, coming to know that the raven is black makes the hypothesis more plausible, and consequently an agent's willingness to accept the hypothesis might increase. This intuition holds for theories considerably more complicated than H. For example, Einstein's General Relativity Theory (GRT), published in 1915, entails that light bends around the sun. This in turn entails that stars must appear further apart during the day than they do during the night to an observer on earth (Figure 1). Arthur Eddington observed in 1919 that this was indeed the case. This was taken to be a major argument in favour of GRT. Philosophers of science, and, more importantly, scientists themselves, would say that this evidence confirms GRT.

One influential systematic account of theory confirmation is Bayesian Confirmation Theory (BCT). One of the charges against this account is that it has the clearly unacceptable consequence that already known evidence can never confirm a newly proposed theory. Bayesianism and the alleged 'problem of old evidence' are the focus of this paper. In *Section I*, I outline Bayesian confirmation theory, and I show how the problem of old evidence arises within it. In *Section II*, I go on to examine two versions of a solution that has been proposed to the problem of old evidence, the counterfactual solution, and show that both are unsatisfactory. In *Section III*, I draw on a proposal by Marc Lange to put forward a different, although related, solution to the problem. To do so, I begin by arguing that the traditional Bayesian framework needs to be modified in order for it to better fit scientific practice, as proposed by Lange. I then show that, in this modified framework, with the addition of a proviso, the problem of old evidence can be solved.

I. The Problem of Old Evidence in Bayesian Confirmation Theory

The Bayesian approach to confirmation relies on the probability calculus, and particularly on a central theorem proven by Bayes that is a direct consequence of the probability axioms:

$$p(A|B) = \frac{p(B|A).p(A)}{p(B)}, \text{ given } p(B) \neq 0.$$

Bayesian epistemology aims at giving an account of how a rational agent ought to update her system of beliefs in the light of new evidence. The standard account is called Bayesian conditioning. According to that model, an epistemic agent starts off with a prior probability distribution p_1, her credences, over a given set of propositions. The requirement that her degrees of belief satisfy the probability calculus embodies the Bayesian constraint of 'coherence'. She then updates that prior distribution by conditioning on the propositions that she learns to obtain a posterior probability distribution p_2. The conditioning consists of using the rules of probability, including notably Bayes' theorem, in order to incorporate the proposition that is learnt into the set of background knowledge in such a way that the agent remains coherent. Formally speaking: a probability function p_2 results from a probability function p_1 by conditioning on a proposition e iff for all propositions a,

$$p_2(a) = p_1(a|e)$$

Effectively, at time t_2, the agent has the proposition e within her background knowledge. If an agent begins her 'epistemic life' with a coherent probability distribution over her beliefs, and then updates by Bayesian conditionalisation, her resulting credence distribution will also satisfy the coherence requirement. Reiteration of the conditionalising process may then occur.

Bayesian confirmation theory uses this epistemological framework to provide a criterion of confirmation. Let $p_1(H|E)$ be called the posterior probability of H, $p_1(H)$ its prior probability, and $p_1(E)$ the prior probability of E. The Bayesian criteria of confirmation are the following: at some time t_i, for any piece of evidence E and any hypothesis or theory H,

- E *confirms* H if and only if $p_i(H|E) > p_i(H)$.
- E *disconfirms* H if and only if $p_i(H|E) < p_i(H)$.
- E is *neutral towards* H if and only if $p_i(H|E) = p_i(H)$.

Intuitively, this is appealing, because, since, by definition, $p_i(H|E) = p_{i+1}(H)$, the above account implies that, if some piece of evidence E confirms H, then $p_{i+1}(H) > p_i(H)$ – the agent's credence in H increases as she learns E. Furthermore, given that, as stated above, if a prior distribution satisfies the coherence requirement, then so does the posterior distribution obtained by Bayesian conditioning, we can introduce new evidence after having already updated, and this opens the possibility of incremental confirmation. For the sake of clarity, the subscripts related to probability functions will be omitted in the rest of the discussion if all the occurrences of p in a formula refer to the same function.

To illustrate how BCT works, let us show how it handles the historical example given in the introduction. Our hypothesis H is GRT, and our evidence E is the result of Eddington's test. We want to know whether E confirms H. As explained in the introduction, H entails E[1] and therefore the likelihood of E, $p(H|E)$, equals 1. Let us assume that, given our background knowledge, $p(E) = 0.1$ and $p(H) = 0.05$.[2] Thus, according to Bayes' theorem,

[1] Of course, H does not entail E on its own, it does so in conjunction with a large set of auxiliary hypotheses, including for example some laws of optics. This is known as Duhem's problem.

[2] One major problem of Bayesian confirmation theory, called the problem of the pri-

$$p(H|E) = \frac{p(E|H).p(H)}{p(E)} = \frac{1 \times 0.05}{0.1} = 0.5.$$

We have $p(H|E) > p(H)$, so E confirms H - as per our intuitions. Bayesian confirmation theory is an attractive account of confirmation, not only because it has such intuitive results in cases like the above, but also because it can give some solutions to several issues in the philosophy of science. For example, it can account for our intuition that if a hypothesis ($p(H) = 0.05$) makes two predictions, one that seems unlikely ($p(E_1) = 0.1$) and the other likely ($p(E_2) = 0.7$), the observation of the former will have a greater confirming impact than the latter. Indeed,

$$p(H|E_1) = \frac{p(H).p(E_1|H)}{p(E_1)} = \frac{0.05 \times 1}{0.1} = 0.5$$

$$p(H|E_2) = \frac{p(H).p(E_2|H)}{p(E_2)} = \frac{0.05 \times 1}{0.7} = 0.07$$

So, the following holds. Both E_1 and E_2 confirm H, but E_1 does so (considerably) more than E_2:

$$p(H) < p(H|E_2) < p(H|E_1)$$

But notwithstanding its successes, BCT has some problems. This paper aims at developing a solution to one of them, called the problem of old evidence, first raised by Clark Glymour.[3] Further to the result of Eddington's test, another piece of evidence is taken to lend very strong support to GRT: the explanation of the 'anomaly' in Mercury's perihelion. According to Newtonian physics, the orbit of a given planet, to a first approximation, traces out an ellipse with the sun at one of its focus. The point where a planet is closest to the sun is called its perihelion (see Figure 2).

A number of factors cause the position of planets' perihelions to shift, the

ors, is how to determine the values of the prior probabilities of E and H. Curd and Cover (2012) give a good review of the debate around this issue. The values that I have assigned to $p(E)$ and $p(H)$ are however not completely arbitrary. Indeed, I could not have assigned a lower prior to E than I have to H, because since H entails E, and $H \neq E$, $p(E) > p(H)$ must hold.

[3] Glymour, 'Why I Am Not a Bayesian' 63-93.

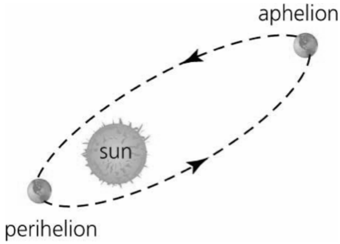

Figure 2: the Perihelion of a Planet

most important of which is the interference with other planets' gravitational fields. Newtonian mechanics, once it took these extra factors into account, succeeded in predicting the positions of all planets' perihelions with remarkable accuracy, except for that of Mercury, which had an anomalous advance of around 43 arcseconds per tropical century.[4] This was acknowledged to be a problem for celestial mechanics in 1859. Many hypotheses were brought forward in order to account for this precession in Mercury's perihelion, including Urbain Le Verrier's postulate of a new planet, Vulcan, between Mercury and the Sun, but all such attempts at solving the puzzle turned out to be unsuccessful. Einstein's GRT, which was constructed for reasons completely independent of this issue, was later discovered to entail exactly an extra advance of 43 arcseconds per tropical century in Mercury's perihelion, as well as the same shifts in its position that Newtonian mechanics predicted. This was taken to lend significant support to GRT.

The problem of old evidence arises in BCT as it is unable to account for this case. The 'anomaly' (it was anomalous within the Newtonian framework) in the perihelion's precession, call it E, was discovered half way through the 19th century. It would seem then that, on any sensible analysis, any rational

[4] An arcsecond is a measure of angle. One arcsecond corresponds to 1/3600 of a degree. A tropical year is the time that the sun takes to return to the same position in the cycle of seasons, as seen from the Earth. One hundred of those compose a tropical century.

agent knew in 1915 that E held and therefore ascribed it a probability equal to 1. Since GRT entails E, we have $p(E|GRT) = 1$. Regardless of the value of $p(GRT)$, the following then holds:

$$p(GRT|E) = \frac{p(E|GRT).p(GRT)}{p(E)} = \frac{1 \times p(GRT)}{1} = p(GRT)$$

This result is more general: if $p(E) = 1$, then for any H, whatever its logical relation to E, the following is true:

$$p(H|E) = \frac{p(E|H).p(H)}{p(E)} = \frac{1.p(H)}{1} = p(H)$$

The reason why, if E is known (ie. $p(E) = 1$), then $p(E|H)$ also equals 1, is that $E|H$ is a subset of E (it is the intersection of E and H), and so $E|H$ is implied by E. Therefore, if one knows E, and is a perfect logician, as is assumed in the Bayesian framework, then one also knows $E|H$.

On BCT, old evidence cannot confirm theories. Intuitively, this is very implausible – it goes against very firm intuitive judgments in particular cases, like the orbit of Mercury, and in general there seems to be no good reason why the fact that a piece of evidence is known should, on its own, rule out that evidence from being confirmatory. The next sections of this discussion will aim at modifying BCT in such a way that it can account for the intuition that E confirms GRT.

II. The Counterfactual Response and Its Problems

Two attempted solutions of the problem of old evidence have been discussed in the literature, namely i) adopting a new criterion for confirmation, or ii) denying that $p(E) = 1$. Many have argued for a solution of the first type, for example by claiming that we should conditionalise on the recognition (which did not of course occur before 1915) that E is implied by H, rather than on the fact that E is true.[5] I focus here on the second type of solution. The most common response of this type is the counterfactual response: in a nutshell,

[5] A review of such proposals can be found in Curd et. al., *Philosophy of Science* 626.

this response claims that we should assign to $p(E)$ the value that it 'would have had, had E not been known.' Martin Curd and J.A. Cover[6] identify two types of counterfactual responses, the historical proposal and the present proposal. I examine them in turn and show that neither works as a response to the problem of old evidence.

a. The 'Historical Proposal'

Glymour, after introducing the problem of old evidence, concedes that a natural response would be to introduce counterfactuals. His suggestion is that we should assign to $p(E)$ the value that it had before we came to know E. He draws an analogy with a coin-flipping situation. Suppose that you toss a coin, and the outcome is heads. The prior probability of the coin to yield outcome heads should be, not 1, but ½ - the probability that was assigned to that outcome before the experiment. Applied to the example of GRT and the precession of Mercury's perihelion, he proposes that one could take the value of $p(E)$ to be that before the precession was observed.

However, Glymour notes two main problems with such a proposal. The first is that there would be a technical difficulty linked to the Bayesian calculus. We cannot simply use the degree of belief in E that we had at a previous time and retain our other present-day degrees of belief in all other propositions, because that would most probably violate the coherence requirement. But even if this difficulty could be overcome, Glymour raises another, more serious philosophical objection to the historical proposal: choosing which historical period is relevant is not as straightforward as it seems. The discovery of the precession in Mercury's perihelion took place in stages between the mid nineteenth century and 1912, with estimated values changing and confidence in the methods of measurement varying. Thus, at the different times between these two dates, the value of $p(E)$ fluctuated. How do we decide which date to choose as the basis for $p(E)$ in our counterfactual approach? It seems that any such decision would be arbitrary.

[6] Curd et. al., 624.

I argue that there is a third, related problem with the historical proposal. Suppose that it were possible to determine which period to choose as a basis for assigning a value to $p(E)$. Even then, this value would be arbitrary in that it would depend on factors unrelated to either the present beliefs of the agent or the evidence itself. Suppose that it had been decided that the appropriate historical period to determine $p(E)$ was in 1619, at the time of the publication of Kepler's three laws of planetary motion. The paradigm in which scientists were operating at the time comprised a claim about the constancy of the area speed of planets. Thus, their credence in E would have been extremely small. It seems arbitrary to use this value for $p(E)$ today, because their reasons for attributing such a small prior to E have nothing to do either with observations of Mercury's orbit, nor with any theories that are accepted today. Had the value of $p(E)$ determined by an agent in 1619 been greatly influenced by observations of the phenomenon, or by hypotheses that are still accepted today as plausible, it could have made some sense to use it as a counterfactual value, but this was not the case. I used a rather extreme example to illustrate my point – it is unlikely that, if a period for determining the prior of E should be used, it would be 1619 – but the fact is that the same criticism can apply to whatever period is chosen, because the accepted hypotheses in science vary greatly with time. The historical proposal being inadequate, I turn to the 'present proposal.'

b. The 'Present Proposal'

The 'present proposal', which is upheld notably by Colin Howson and Peter Urbach,[7] consists in calculating the prior of E with respect to all present beliefs except E. There are two steps: the first is to 'delete' E from the background knowledge B, and the second is to calculate $p(E)$ given the new background knowledge B\E. This proposal is not vulnerable to any of the criticisms against the historical proposal, because there is no talk of historical period. However, it has some problems of its own.

[7] Howson and Urbach, *Scientific Reasoning*

The first is brought forward by Howson and Urbach themselves. The issue is that the content of B\E depends on the way that B and E are formulated. Suppose that an agent's background knowledge contained the propositions a and b. Because Bayesian agents are assumed to be perfect logicians, the proposition $a\&b$ is also part of the agent's background knowledge. Simply removing a from B thus makes little sense, unless it is removed from an axiomatisation of B, in which case all occurrences of a would be removed. However, there are many different equivalent axiomatisations of B. For example, even in a simple case where intuitively all we know is a and b, B could be axiomatised as either $\{a,b\}$ or $\{a,a \rightarrow b\}$. But removing a from these sets leaves us with two different sets: $\{b\}$ and $\{a \rightarrow b\}$ – and there seems to be no non-arbitrary way of deciding which one to use between the two.

Even if this problem could be solved (Howson and Urbach point to a paper by Satoru Suzuki[8] which claims to do so), there still remains a problem, brought forward by Salmon.[9] It is that of calculating the prior of E, $p_c(E)$,[10] in terms of B\E. The suggestion brought forward by Howson and Urbach, which is rather standard, is that, since $p_c(E) = \sum_{i=1}^{k} p_c(E|H_i).p_c(H_i)$ holds where H_i is the set of all alternatives to H including H, it suffices to calculate $\sum_{i=1}^{k} p_c(E|H_i).p_c(H_i)$. That means, for all alternatives $H_1, H_2, \ldots, H_k$, that:

$$p_c(E) = p_c(E|H_1).p_c(H_1) + p_c(E|H_2).p_c(H_2) + ... + p_c(E|H_k).p_c(H_k)$$
$$+ p_c(\neg(H_1 \vee H_2 \vee \ldots \vee H_k)).p_c(E|\neg(H_1 \vee H_2 \vee \ldots \vee H_k))$$

But, as Salmon points out, the last term is intractable. This is because, even if all the alternative hypotheses considered by an agent are mutually

---

8 Suzuki, 'The Old Evidence Problem and AGM Theory' 105-126.

9 Curd et. al., 624.

10 The reader is reminded that, in this subsection, $p(E)$ is the agent's counterfactual credence in E, as if E was not known. To make this clear, I will use the denotation p_c to refer to this function throughout the subsection. However, keep in mind that the fact that we are concerned with counterfactual values only affects the value of $p(E)$, and the values of $p(E|x)$ for all x. That is, $p_c(E)$ differs from $p_{real}(E)$ and $p_c(E|x)$ from $p_{real}(E|x)$, but $p_c(x)$ does not differ from $p_{real}(x)$ for all other x.

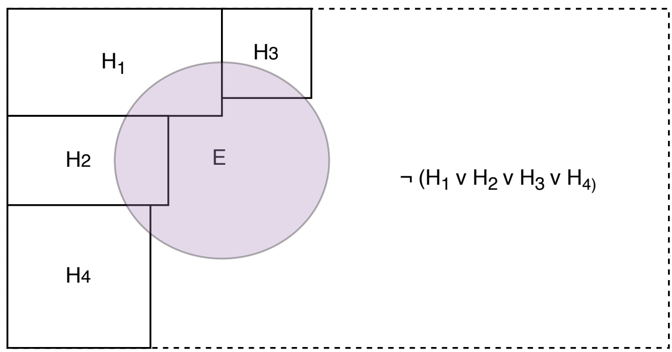

Figure 3: Intractability of $p_c(E)$

exclusive, they will never exhaust the possible hypotheses to explain a phenomenon, and most importantly, it will always be impossible for that agent to know how big a proportion of possible explanations he has taken into account. Let us illustrate this on a diagram. Suppose that an agent can think of four mutually exclusive hypotheses H_1, H_2, H_3, H_4, represented below as rectangles whose areas represent the credence that the agent has in these hypotheses (Figure 3). She has no way of knowing how large the area representing the possibilities that she has not thought of, exactly because she has not thought of them. Thus, according to Salmon, even if there was a way to delete E from B, it would still be impossible to deduce $p_c(E)$ from it.

However, Howson and Urbach object to this response by recalling that the Bayesian framework interprets probabilities in a subjective manner, probabilities are degrees of belief. Thus, since the agent does not take $\neg(H_1 \vee H_2 \vee \ldots \vee H_k)$ into account in her system of beliefs (she has not considered it and so her credence in $\neg(H_1 \vee H_2 \vee \ldots \vee H_k)$ is zero), the above diagram is not an adequate representation of the problem, and the dotted line and everything in it should be deleted. Thus, that last term may be discarded and $p_c(E)$ becomes:

$$p_c(E) = p_c(E|H_1).p_c(H_1) + p_c(E|H_2).p_c(H_2) + ... + p_c(E|H_k).p_c(H_k)$$

This is no longer intractable. To illustrate their point, they explain that, in 1915, the only serious alternative to GRT was Classical Gravitation Theory (CGT), which was part of Newtonian Mechanics. Thus, according to them, the probability of the precession in Mercury's perihelion was:

$$p_c(E) = p_c(E|CGT).p_c(CGT) + p_c(E|GRT).p_c(GRT)$$

Therefore, Salmon's criticism fails to show that it is impossible, if E can be 'removed' from B, to calculate the prior probability of E.

I argue however that, although the calculation is possible, it leads to a conceptual difficulty linked to the standard Bayesian framework. The 'present proposal' requires that, at some time t, the value of $p(E)$ be taken to be less than 1, when it had been fixed at 1 at a previous time t-1. But this deletion of knowledge is impossible in the standard Bayesian framework which works as a series of updates by Bayesian conditionalisation: the calculus does not allow for an agent to lower her probability from 1 to a lesser value. Furthermore, the traditional model does not allow for an agent's beliefs (even if counterfactual) not to correspond to a point in the Bayesian updating. The 'present' counterfactual move is therefore inconsistent with the assumptions of the traditional Bayesian framework. What options do we have to go forward? On the one hand, we could reject the present counterfactual move altogether. On the other, we could modify the traditional Bayesian framework, so as to make it consistent with a solution to the problem of old evidence that, just like the present proposal, is based on the idea that there is a reasonable prior for E that does not construe it as old. In *Section III*, I opt for the latter strategy.

III. Rational Reconstruction versus Updating: Towards a Satisfactory Solution

In this third section, I argue that, if we enlarge the scope of traditional Bayesianism in the way suggested by Mark Lange,[11] the present proposal can be modified so as to provide a satisfactory answer to the problem of old evidence.

[11] Lange, 'Calibration and the Epistemological Role of Bayesian Conditionalization.' 294-324.

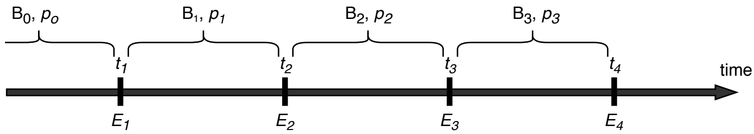

Figure 4: Bayesian Updating

First, I explain how Lange differentiates between updating and justification, and how he uses that distinction to expand the standard Bayesian framework into another that is richer and that coincides better with our intuitions about confirmation. Then, I outline how Lange claims to use this extended framework to solve the problem of old evidence, I show some limitations to his proposal and suggest a way in which to build on it in such a way that it becomes satisfactory.

a. Justification versus Updating, an Extended Bayesian Framework

In this subsection, I present and explain Lange's suggestion of differentiating *update* and *justification* by building a new Bayesian framework that incorporates both the diachronic and the synchronic aspects of Bayesian epistemology. In order to illustrate the arguments I make in this subsection, I introduce an example. Alice presents some symptoms of tuberculosis and her doctor, Bob, wants to know whether she has contracted the illness. He performs a series of tests on Alice. In order for tuberculosis to be diagnosed, three tests must come back positive.

In the previous section, I explained how the process of Bayesian updating via Bayesian conditionalisation works: the epistemic agent begins with a prior probability function over the set of her beliefs, and this function is updated according to Bayesian conditionalisation at each step at which she encounters new evidence (where each E_i is taken to be the strongest evidential claim she learns at that stage). This can be represented as on Figure 4.

Here, the sets B_i are the sets of beliefs of the agent, and the probabilistic functions p_i are her credence distributions over B_i. At time t_1, the agent learns evidence E_1 and updates her credence distribution, so that it remains consistent as it changes from p_0 to p_1. If we apply this to the example, we get the following. Bob, before the first test, has credence say 0.3 in the claim that Alice has contracted tuberculosis, in view of the symptoms. He then, at t_1, gets the results of the first test, incorporates this evidence into his set of beliefs, and updates his credence in Alice having tuberculosis using Bayes' theorem to, say 0.5. This process is repeated at times t_1, t_2, and t_3, until his credence in that proposition becomes 1 (or very close to 1) at time t_4, supposing that this is when he gets the result of the third test.

A 'Bayesian rational reconstruction' differs from such updates. Indeed, unlike the latter, a justificatory argument is synchronic. It consists, as Lange explains, of several steps. It begins with a primary probability distribution over a chosen set of beliefs. The person justifying her credence in a proposition chooses which beliefs she would like to take as 'primary' – this is the background knowledge that she considers relevant (B'_0 in the figure 5 below) – and she assigns a coherent probability distribution over them (p'_0). Then, at each step s_i, she brings in new evidence E_i and 'updates' her credence function over her set of beliefs according to Bayesian conditionalisation. She then ends up with B'_{i+1} and p'_{i+1}, and can reiterate the process. Lange himself raises the obvious issue of how the agent should determine her initial, or primary, set of beliefs B'_0, and the associated probability function p'_0. He responds that this is no more problematic than the problem of the priors in standard Bayesian confirmation theory. This is, I believe, not a satisfactory response. Of course, just like the prior distribution at the beginning of an agent's epistemic life (supposedly then, at birth), the primary distribution in a justificatory argument is subject to the problem of the priors. But one of the widely accepted arguments against the significance of that problem for standard Bayesian epistemology is the washing out of the priors – namely the fact that, as evidence accumulates, the value of the prior influences that of the posterior less – cannot be transposed to the case of justificatory arguments, because justifications are much shorter than epistemic histories, and therefore evidence cannot accumulate enough. The Bayesian who accepts this expanded framework will need to provide an additional proviso on the primary probability distribution of justifications. I will propose one later in the paper.

There are two things to note about the relation between updates and justificatory arguments. The first is that, in a justificatory argument, the person justifying her belief can choose which evidence to incorporate in her argument. In the diagram above, for example, the agent has omitted E_3. The second thing to note is that the set of beliefs B'_3 that the agent has at the end of her justificatory argument is a subset of the set of beliefs that the agent really has after t_4, B_4. Indeed, B_4 is the set containing all of the agent's beliefs, and B'_3 is only one or few of such beliefs. Let us now illustrate this Bayesian justificatory argument using the example of Alice. Bob justifies at t_4 his belief that Alice has tuberculosis. He begins by claiming that his original set of relevant beliefs B'_0 contains the belief that Alice has tuberculosis and he assigns a probability 0.3 to this proposition. Then, he brings in the result of the first test, E_1, incorporates it into his justificatory beliefs and updates his credence distribution – so that his credence in Alice being ill becomes 0.5. Suppose that E_3 was the name of Alice's mother. Learning that evidence, although it updated some of Bob's beliefs, had no incidence on his belief that Alice had tuberculosis. Therefore, it is reasonable for him not to bring it in as evidence in the justificatory argument. He ends up with a set of justificatory beliefs B'_3 (here, that Alice has contracted tuberculosis) which is a subset, here proper, of the set B_4 of all his beliefs.

b. Solving the Problem of Old Evidence

When you ask whether observing a black raven confirms the hypothesis that all ravens are black, you really could be asking two questions. The first is: as you *discovered* that an object was a black raven, did your credence in the hypothesis increase? This is a question about update, and it is a *factual* and *historical* question: the agent is asked whether the update took place when she learnt the new evidence. The second question differs: can you give a

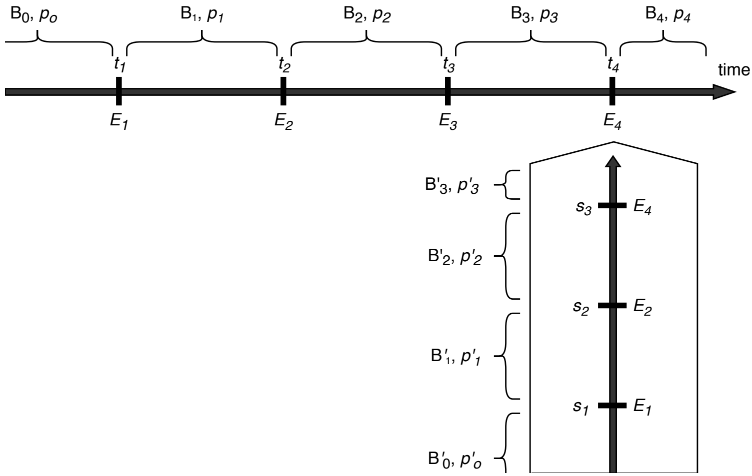

Figure 5: Bayesian Updating and Justification

satisfactory[12] *justification* for your degree of belief in a hypothesis, within which your degree of belief increases from $p(H)$ to a higher $p(H|E)$? Thus, in the new framework, there are really *two* questions of whether old evidence can confirm a hypothesis. The first is whether this can happen in the update, that is, in the real epistemic history of an agent, and the second is whether an agent can give a justification that old evidence confirms a hypothesis. In this last subsection of the paper, I show how the introduction of the new framework which contains both updates and justifications leads to the dissolution of both problems of old evidence. First, I uphold that in the updating, there can indeed be no confirmation of a new theory by old evidence, and explain why this might seem counterintuitive though in fact, it is not. Second, I show how, by a move similar to the counterfactual present proposal, old evidence can confirm a hypothesis in a justificatory context. Finally, I discuss some problems, including those that applied to the counterfactual solutions (both historical and present), and show that none of threaten the proposed solution.

[12] I will discuss what makes for a satisfactory justification at a later stage.

In a context of Bayesian *updating*, there can be no confirmation by old evidence. I explained why this was the case above: if the prior probability of E at the time H is proposed is 1, then the posterior probability of H, $p(H|E)$, is equal to its prior, $p(H)$. As stated above, this might seem very counterintuitive, because it seems that in many examples, such as that of Einstein's GRT and the precession of Mercury's perihelion, old evidence *can* confirm new hypotheses. But I think that this is only counterintuitive because there is a conflation between the two types (updating and justificatory) of confirmatory arguments: we can justify (I will explain how in the next paragraph) that the evidence from Mercury can confirm GRT, and therefore think there must also be confirmation in the update. But actually, it fits our intuitions to deny this latter claim. Indeed, there is no moment in the history of the agent's 'epistemic life' where the probability of H is increased by conditionalising on E. However, although there is no *confirmation*, the fact that E was already known has an *impact* on the probability H is assigned when it is proposed. Remember that the whole framework assumes that the agent is a perfect logician and that her credences respect the coherence requirement. Suppose that the agent already knew E, her credence in that evidence is 1. As she was introduced to the hypothesis H, she had to assign a probability value to H, which had to be consistent with the rest of her beliefs. Since she already knew E, and since she is a perfect logician so she knows that it is entailed by H, she must assign a relatively high prior probability to H in order to be coherent. Thus, the prior probability of H, $p(H)$, is affected by the fact that the agent knew E. Had she not known E, the probability of H might not have been so high because the coherence requirement would not have demanded it. Thus, although there is no confirmation in the *update*, E has an impact on the prior of H – and this corresponds to intuitions.

There is however confirmation in the *rational reconstruction*. As Lange claims, the agent can construct a justificatory argument in which E counts as confirming evidence for H. The way that she does this is rather similar to the earlier 'present proposal'. She takes as her primary set of relevant beliefs a subset of her actual set of beliefs which contains the hypothesis H and all auxiliary hypotheses which she accepts do support H. Over this set, she builds a coherent credence distribution such that her degree of belief in E is less than 1. She then uses Howson and Urbach's proposal to determine the value of $p(E)$, given that primary set of beliefs. This is similar to the present proposal of the counterfactual response in that the agent has, at the introduction of her justificatory argument, a coherent probability distribution over a set of beliefs which corresponds to what she would know that is relevant to H if she didn't know E. Once the agent has fully specified this primary state, at stage 1, she introduces the evidence E as known – $p(E) = 1$. She must then update her original credence distribution by Bayesian conditionalisation, so as to be consistent with this 'new' evidence. In doing so, it will presumably be the case that $p(H|E) > p(H)$, and thus E confirms H in the context of a justificatory argument. The old evidence that Mercury's perihelion is precessive can still confirm GRT for a contemporary agent: she need only give a rational reconstruction of how she arrived at her current high degree of belief in GRT, introducing the evidence about Mercury's orbit at some non-primary state of the argument.

Let us consider an important issue with Lange's suggestion. It might seem that, using such rational reconstructions, any evidence can be made to confirm any hypothesis. For example, an agent could start a justification of the Ptolemaic model of celestial movement, without construing the data of planetary motion in the primary set of beliefs as known. One could then introduce it at some stage of the argument, in which case it would count as confirming evidence. But intuitively, data from the stars does not confirm the Ptolemaic model, because the parameters of that model were constructed exactly to fit the existing data. This issue arises because there are absolutely no constraints on what the primary probability distribution in a justification should be. As mentioned before, Lange says about them that they need to be 'unbiased' and 'impartial', but that any further discussion would be an attempt to solve the problem of the priors, which he does not intend to do.[13] It is indeed true that *fully* resolving this issue would amount to solving the problem of the priors,

[13] Lange, 303.

but I nevertheless believe that, although a full resolution is complicated to say the least, some improvement in the case of rational reconstruction is possible. Although, as we have seen already, the justification of a belief need not be identical to the historical arrival to that belief, there is a case for arguing that it should be a possible way of actually arriving at that belief using Bayesian updating. Given an initial credence distribution (here clearly the original problem of the priors applies), there are a number of different orders in which the evidence that was acquired could have been acquired. In my example above, it could have been that Alice's second test results came out before the first ones, and thus Bob would have acquired his final credence in Alice's having tuberculosis in a different way. I argue that the steps of a justification should be one such *possible* order, and that this suitably restricts the possible primary distributions. There is something about justifications, or *rational* reconstructions, that gives them a particular epistemic status, and this is that, for them to be satisfactory, they need to be anchored in reality in some way. A justification of Alice's having tuberculosis that would appeal to her mother's name, for example, would be unsatisfactory. The reason why this might be is that, during Bob's actual Bayesian updating, the name of Alice's mother did not have an impact on his credence in Alice having tuberculosis, and it could not have because the two are unrelated. Given this remark, I propose the following proviso: it must be the case for a justification to be satisfactory that each step of the justification, including the primary stage, *could have been* a step in the actual update, given the agent's actual initial credence distribution at t_0. Furthermore, the more *plausible* the justification, the more satisfactory it is. Plausibility here is to be understood as degree of closeness to reality, and it could be measured probabilistically: a justification would be plausible at degree 1 if it was identical to the actual update, and 0 if it was logically impossible for it to have been an actual update.

I argue that this entails a probabilistic version of Elie Zahar[14] and John Worrall's[15] criterion for confirmation by old evidence. Their original criterion

[14] Zahar, 'Why Did Copernicus's Research Programme Supersede Ptolemy's?'
[15] Worrall, 'New Evidence for Old'

states that evidence must not have been used in the construction of the theory for it to confirm that theory. Above, I mentioned that using this framework, it seems that the Ptolemaic model in 1543 can be confirmed by planetary data: the agent could simply introduce the theory in the primary stage of her justification and the data at subsequent steps. But, given the proviso from the above paragraph, such a justification would be extremely unsatisfactory. This is because the intricacy of the Ptolemaic model makes it extremely unlikely for someone to conceive of it without knowledge of the data, and according to the proviso, a justification based on such unlikely facts would have very a very low plausibility. This can be understood as a probabilistic version of the Zahar-Worrall criterion: if the evidence used in the construction of the theory is introduced after the theory in a justificatory argument, because this is so unlikely to have happened in reality, the plausibility (which, as explained in the previous paragraph, can be understood probabilistically), and therefore, the satisfactoriness of the confirmatory justification is very low. Therefore, if a *good* justification must be a *plausible* possible update, the only admissible 'primary' degree of belief in any E that was used in the construction of some theory H is 1.

This proposal is not vulnerable to any of the criticisms brought forward against the counterfactual response. All the criticisms against the 'historical proposal' dissolve immediately as there is no reference to any historical period in a justificatory argument. This proposal is also immune to the criticisms against the 'present proposal'. As we saw in Section II.a., the two serious objections are that there is no non-arbitrary way of deleting E from background knowledge, and that even if there were, it would be inconsistent with the traditional update-only Bayesian framework to do so. In the update+justification framework, unlike in the update-only framework, there is no need to delete E from background knowledge, there is only a need to construct a primary credence distribution such that the credence in E is not 1. Thus, the agent might choose the primary credence distribution that she pleases – and here the only remaining problem is that of the priors. The update+justification framework also resolves the issue of the internal inconsistency related to the deletion of

E from background knowledge. Indeed, in that framework, the agent can simply fail to include E in the primary set of beliefs of her justification – provided it respects the proviso. This is perfectly coherent with the general framework.

Before concluding, I consider a potential problem for using justification as a basis for confirmation. It is possible to construct two justifications for a belief, one in which there is confirmation, and another in which there is not. For example, Bob could give two justifications of his belief in Alice having tuberculosis: one in which he begins by introducing the results of the three tests and then formulates the hypothesis that Alice is ill, and another in which he firsts introduces his hypothesis and then brings in the evidence. There would therefore be confirmation of the hypothesis by the evidence only in the second justification. Lange does not consider that the same belief could be justified by different rational reconstructions, and thus cannot directly deal with such a problem. However, once the proviso I put forward is added to Lange's proposal, this problem can be solved. One of those two justifications is more satisfactory because it differs less from reality, here, the second one. Because, in the better justification, there is confirmation, that is the result we should consider. Can such a move really be satisfactory? In historical cases of old evidence, like GRT and the advance in Mercury's perihelion, the best justification according to my proviso is the one that is closest to reality, namely the one that begins by introducing the evidence and then introduces the theory. Thus, in the best justification, there is no confirmation. Is that problematic? As I have explained above, it is actually not counterintuitive to think that there is no actual confirmation in such cases. Furthermore, there is an almost imperceptibly less plausible (that means, extremely plausible since the one that copies reality has a plausibility of 1) justification in which confirmation does occur. I think this accounts for our intuitions about cases of old evidence.

Conclusion

In this paper, I have sought to present a solution to the Bayesian problem of old evidence, building on a suggestion by Marc Lange. To do so, I began, in the first section, by outlining the basic tenets of Bayesian confirmation theory, and by showing how the problem of old evidence arises within it. In the second section, I went on to examine the counterfactual response to this problem, by investigating the historical and present counterfactual proposals. I concluded that neither of them is satisfactory. Finally, in the third section, I drew on Lange's suggestion of a Bayesian rational reconstruction distinct from updating to elaborate a new more intuitively appealing Bayesian framework comprising both aspects, and I showed that, in this framework and with a proviso, the problem of old evidence could be solved in a way that is immune to the criticisms brought forward against the counterfactual response.

References

[1] Curd, M., J.A. Cover, and C. Pincock. *Philosophy of Science: The Central Issues.* Second International Student Edition. New York: W. W. Norton & Company, 2012.

[2] Glymour, C. 'Why I am not a Bayesian'. *Theory and Evidence.* Chicago: University of Chicago Press, 1981. 63-93.

[3] Howson, C., and P. Urbach. *Scientific Reasoning: The Bayesian Approach.* Open Court Publishing, 2006.

[4] Lange, M. 'Calibration and the Epistemological Role of Bayesian Conditionalization.' *Journal of Philosophy* 96, no. 6 (1999): 294–324.

[5] Suzuki, S. 'The Old Evidence Problem and AGM Theory.' *Annals of the Japan Association for Philosophy of Science* 13, no. 2 (2005): 105–26.

[6] Worrall, J. 'New Evidence for Old.' In *In the Scope of Logic, Methodology and Philosophy of Science*, Eds. Peter Gardenfors, Jan Wolenski and Katarzyna Kijania-Placek., 191-209. London: Kluwer Academic, 2002.

[7] Zahar, E. 'Why Did Copernicus's Research Programme Supersede Ptolemy's?' In Lakatos, I., *The Methodology of Scientific Research Programmes*, Vol. 1. Cambridge University Press, 1978.

Chloé de Canson has just graduated from LSE with a BSc Philosophy, Logic and Scientific Method (2012 - 2015). From October 2015, she will be reading the MPhil Philosophy at the University of Cambridge. Her main areas of interest are in philosophy of science, philosophy of probability and logic. You can contact her at [chloe.decanson@gmail.com].