A New Solution to the Problem of Old Evidence

Stephan Hartmann*

July 20, 2014

Abstract

The Problem of Old Evidence has troubled Bayesians ever since Clark Glymour first presented it in 1980. Several solutions have been proposed, but all of them have drawbacks and none of them is considered to be the definite solution. In this article, I propose a new solution which combines several old ideas with a new one. It circumvents the crucial omniscience problem in an elegant way and leads to a considerable confirmation of the hypothesis in question.

1 Introduction

The Problem of Old Evidence is easy to state. If the probability of the evidence, $P(\mathrm{E})$, is 1, then the likelihood $P(\mathrm{E}|\mathrm{H})$ is also 1, and hence

$$P^*(\mathrm{H}) := P(\mathrm{H}|\mathrm{E}) = \frac{P(\mathrm{E}|\mathrm{H})\, P(\mathrm{H})}{P(\mathrm{E})} = P(\mathrm{H})\,. \tag{1}$$

Consequently E does not confirm H according to standard Bayesian Confirmation Theory, i.e. if conditionalization is used to compute the posterior probability. This observation conflicts with the practice of science, as Glymour (1980) has forcefully pointed out.[1] Note that using conditionalization here is somewhat dubious as, in this case, nothing new is learned. So why should one even conditionalize?

In this article, I argue that something new happens in the course of the deliberation. The basic idea is this: Once a scientist becomes aware of the logical fact that the hypothesis under consideration entails the evidence (and that other available

*Munich Center for Mathematical Philosophy, LMU Munich, Geschwister-Scholl-Platz 1, 80539 Munich (Germany) – http://www.stephanhartmann.org – s.hartmann@lmu.de.

[1]See Earman (1992: ch. 5) for a critical discussion of a number of well-known responses to the Problem of Old Evidence.

theories do not entail E), she changes her belief about the disjunctive proposition A: "There is an alternative theory that entails the evidence or the evidence is the result of a chance mechanism." This, in turn, prompts an increase of the posterior probability of the hypothesis. Note that we are assuming that the scientist has *beliefs about the origin of the evidence*: E is either a deductive consequence of H, or it is a deductive consequence of an alternative to H, or it is the result of a chance mechanism, where the first and second disjunct are not mutually exclusive. Working out this idea in detail requires six elements:

(i) The scientist's beliefs are always consistent with the claim that she knows that the hypothesis H entails the evidence E. (This circumvents the crucial omniscience problem.) Initially, however, she is not aware of this logical fact.

(ii) Only after doing the deductions, the scientist becomes aware of the logical fact that H entails E (and that other available theories do not entail E). As a result of this,

(iii) the scientist lowers the probability of the disjunctive proposition A. Before the scientist becomes aware of the logical fact that H entails E, the scientist considered it to be quite likely that H does not entail E.[2] Instead she considers it to be quite likely that a (so far unknown) alternative to H entails E (and for which E is evidence) or that E is the result of a chance mechanism. Note that after becoming aware of the fact that H entails E (and that other available theories do not entail E), it is still possible that an alternative to H also entails E. In fact, E could be entailed by any given number of theories.

(iv) The probability of the evidence is $1 - \epsilon$, and not 1 (Fitelson 2004). This reflects the fact that E is a contingent proposition. Note, though, that ϵ can be arbitrarily small, and so this is not a strong restriction for all practical purposes. In fact we will calculate the limit $\epsilon \to 0$ at the end and it will turn out to be finite.

(v) The probability of the evidence does not change in the course of deliberation. More specifically, it is not affected by our deliberations about H or by the possible existence of an alternative to H. This poses a constraint on the likelihoods before and after becoming aware of the logical fact that H entails E (and that other available theories do not entail E).

[2] One could also argue that the scientist should be *agnostic* about the deductive relationship between H and E before she becomes aware of the logical fact that H entails E. We do not consider this possibility here because there is no agreement on how to model ignorance in a Bayesian framework. Cf. Norton (2008, 2011).

(vi) The scientist determines the posterior probability distribution by minimizing the Kullback-Leibler divergence between the posterior probability distribution and the prior probability distribution. Note that this procedure is more general than conditionalization: It leads to standard conditonalization and Jeffrey conditionalization if corresponding constraints concerning the posterior probability of the evidence are added (see Diaconis and Zabell 1982). More recently, Hartmann and Rafiee Rad (2014) have argued that minimizing the Kullback-Leibler divergence can also be used to model the learning of an indicative conditional and that standard objections or problems such as the Judy Benjamin Problem can be rebutted provided that the causal structure of the problem at hand is properly taken into account. Encouraged by these success stories, we conjecture that minimizing the Kullback-Leibler divergence between the posterior and the prior probability distribution is an essential ingredient in any general Bayesian account of belief change.

Note that on our proposal one does not have to conditionalize on E (which arguably does not make sense as the scientist does not learn E) or on the new proposition (after extending the language appropriately) that the hypothesis entails the evidence (Garber 1983). It also avoids counterfactual reasoning as in Howson's solution to the Problem of Old Evidence (Howson 1991).

The remainder of this article is organized as follows. Section 2 presents our general model which works for all cases where the probability of the evidence does not change in the course of deliberation. Section 3 then considers the special case of old evidence where the probability of the evidence is (close to) 1. We conclude, in Section 4, with a short remark concerning the adequacy of the proposed solution.

2 The General Model

We introduce the two usual binary propositional variables.[3] The variable H has the values H: "The hypothesis is true", and ¬H: "The hypothesis is false". The variable E has the values E: "The evidence obtains", and ¬E: "The evidence does not obtain". We consider the case where it is a matter of fact that H entails E. Additionally, we introduce the binary propositional variable A_T which has the values $\mathrm{A_T}$: "There is an alternative hypothesis that entails E", and $\neg\mathrm{A_T}$: "There is no alternative hypothesis that entails E" and the binary propositional variable A_C which has the values $\mathrm{A_C}$: "E is the result of a chance mechanism", and $\neg\mathrm{A_C}$:

[3]Throughout this article we follow the convention, adopted e.g. in Bovens and Hartmann (2003), that propositional variables are printed in (upper case) italic script, and that the instantiations of these variables are printed in (upper case) roman script. Bovens and Hartmann (2003) also introduce the bits of the theory of Bayesian Networks that we use in this article.

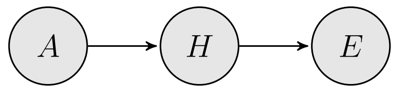

Figure 1: The Bayesian Network for the Problem of Old Evidence.

"E is not the result of a chance mechanism". Let A:= $A_T \vee A_C$. We assume that the scientist has beliefs about A, E and H and that it is a logical fact that H entails E.

The Bayesian Network in Figure 1 encodes the probabilistic dependencies and independencies between the three variables: The variable A directly influences the variable H, the variable H directly influences the variable E, and A influences E only through H: If the hypothesis is true (false), then our credence in E should be equal to the likelihood $P(\mathrm{E}|\mathrm{H})$ (or $P(\mathrm{E}|\neg\mathrm{H})$, respectively).

To complete the Bayesian Network, we have to fix the prior probability of the root node A and the conditional probabilities of all other nodes, given the values of their parents. First, we set

$$P(\mathrm{A}) = a \tag{2}$$

and assume that $a \in (0, 1)$. The value of a will, in fact, be fairly large as the scientist is not yet aware of the logical fact that H entails E. And so she strongly believes that there is an alternative to H that entails E or that E is the result of a chance mechanism.

Second, we set

$$P(\mathrm{H}|\mathrm{A}) = \alpha \quad , \quad P(\mathrm{H}|\neg\mathrm{A}) = 1\,, \tag{3}$$

with $\alpha \in (0, 1)$. If A is false, i.e. if there is no alternative to H that accounts for E and if E is not the result of a chance mechanism, then the probability of H is one. However, if A_T (and therefore A) is true and there is an alternative to H that entails E, then it is possible that H is true and entails E as well because there may be several theories that entail E. Hence $P(\mathrm{H}|\mathrm{A}) =: \alpha > 0$, where α measures how strongly the scientist believes in H, even if there is an alternative theory that entails E. It is important to note that $\alpha > 0$ implies that the propositions A and H are not mutually exclusive.

From eqs. (2) and (3) we calculate the prior probability of the hypothesis using the Law of Total Probability:

$$\begin{aligned} P(\mathrm{H}) &= P(\mathrm{H}|\mathrm{A})\,P(\mathrm{A}) + P(\mathrm{H}|\neg\mathrm{A})\,P(\neg\mathrm{A}) \\ &= \alpha\,a + \overline{a} \end{aligned} \tag{4}$$

From this equation it follows that α is fairly small as a is fairly large and $P(\mathrm{H})$ is not very large. Typical values could be $a = .8$ and $\alpha = .2$, in which case we get $P(\mathrm{H}) = .36$, but our argument of course does not depend on these assignments. Finally, we set

$$P(\mathrm{E}|\mathrm{H}) = 1 \quad , \quad P(\mathrm{E}|\neg\mathrm{H}) = q \,, \tag{5}$$

with $q \in (0,1)$. Here we assume that the beliefs of the scientist are consistent with the logical fact that H entails E.[4] She therefore assigns the conditional probability $P(\mathrm{E}|\mathrm{H})$ the value 1 which makes sure that her beliefs are coherent.[5] Setting $P(\mathrm{E}|\mathrm{H}) = 1$ may be ad hoc for $P(\mathrm{E}) \ll 1$ as the scientist is not yet aware of the logical fact that H entails E. However, if $P(\mathrm{E})$ is close to 1 (which is the case in the Problem of Old Evidence), then setting $P(\mathrm{E}|\mathrm{H})$ to 1 (or to a value close to 1) is a consequence of having coherent beliefs as we will show in Section 3.
The prior probability distribution over A, H and E is then given by

$$\begin{aligned} P(\mathrm{A},\mathrm{H},\mathrm{E}) = \alpha\, a \qquad &, \qquad P(\mathrm{A},\neg\mathrm{H},\mathrm{E}) = \overline{\alpha}\, a\, q \\ P(\mathrm{A},\neg\mathrm{H},\neg\mathrm{E}) = \overline{\alpha}\, a\, \overline{q} \qquad &, \qquad P(\neg\mathrm{A},\mathrm{H},\mathrm{E}) = \overline{a}\,. \end{aligned} \tag{6}$$

For all other instantiations of A, H and E, the prior probability vanishes. Here we have used the convenient shorthand $\overline{x} := 1 - x$, which we will use throughout this article. Here and in the remainder we also use the shorthand notation $P(\mathrm{A},\mathrm{H},\mathrm{E})$ for $P(\mathrm{A} \wedge \mathrm{H} \wedge \mathrm{E})$.
With this, we calculate

$$\begin{aligned} P(\mathrm{E}) &= \alpha\, a + \overline{\alpha}\, a\, q + \overline{a} \\ &= 1 - \overline{\alpha}\, a\, \overline{q}\,. \end{aligned} \tag{7}$$

Next, the scientist becomes aware of the logical fact that H entails E (and that other available theories do not entail E), which prompts her to change her beliefs. More specifically, she reduces the probability of A and sets

$$P'(\mathrm{A}) = a' < a, \tag{8}$$

where P' is the posterior probability measure. Given that the scientist is now aware of the logical fact that H entails E, she does not consider it so probable anymore that there is an alternative theory that entails E or that E is the result of a chance mechanism. Before becoming aware of the logical fact that H entails E, she strongly believed that there is an alternative to H that entails E or that

[4] Note that it is perfectly fine to set $P(\mathrm{E}|\mathrm{H}) = 1$ even if $P(\mathrm{E}) = 1 - \epsilon$. We might be uncertain about whether E and/or H obtain, but we may nevertheless be convinced that E is always true if H is true.

[5] Note that we are only considering deterministic theories here. For indeterministic theories, $P(\mathrm{E}|\mathrm{H}) < 1$ and our model does not apply in the present form. We leave the confirmation-theoretical analysis of old evidence for an indeterministic theory for another occasion.

E is the result of a chance mechanism (although her degrees of belief have always been consistent with the logical fact that H entails E). After becoming aware of the logical fact that H entails E (and that other available theories do not entail E), she still deems it possible that there is an alternative to H that entails E, but she does not believe in it so strongly anymore.[6]

Note that the belief change expressed in eq. (8) does not result from conditionalization. It is more similar to what actually happens in (Jeffrey) conditionalization when the probability of the evidence suddenly changes from one value to another. In the case of (Jeffrey) conditionalization, this change is prompted by an experience. In the case considered in this article, it is prompted by the insight that H entails E (and that other available theories do not entail E).

As H entails E, the scientist also sets

$$P'(\mathrm{E}|\mathrm{H}) = 1\,, \tag{9}$$

which is in line with the corresponding assignment in the prior probability distribution (see eq. (5)).

To proceed, we assume that the Bayesian Network depicted in Figure 1 remains unchanged after the agent changed her beliefs about A. Hence, the posterior probability distribution has the following form:

$$\begin{aligned} P'(\mathrm{A},\mathrm{H},\mathrm{E}) &= \alpha' a' \quad , \quad & P'(\mathrm{A},\neg\mathrm{H},\mathrm{E}) &= \overline{\alpha'}\, a'\, q' \\ P'(\mathrm{A},\neg\mathrm{H},\neg\mathrm{E}) &= \overline{\alpha'}\, a'\, \overline{q'} \quad , \quad & P'(\neg\mathrm{A},\mathrm{H},\mathrm{E}) &= \overline{a'}\,, \end{aligned} \tag{10}$$

where we have replaced all variables with the corresponding primed variables. Note that the value of a' is already fixed (as the scientist sets it to a lower value after becoming aware of the logical fact that H entails E and that other available theories do not entail E)[7], but the values of α' and q' have to be determined. We do this by minimizing the Kullback-Leibler divergence between P' and P taking all relevant constraints on P' into account.

Here are two additional constraints on the posterior probability distribution (besides eq. (8)). First, we assume that the probability of E remains unchanged in the course of deliberation, i.e.

$$P'(\mathrm{E}) = 1 - \overline{\alpha'}\, a'\, \overline{q'} \equiv P(\mathrm{E}) = 1 - \overline{\alpha}\, a\, \overline{q}\,. \tag{11}$$

[6]It could be objected that a' should be greater than a because it is easy to construct alternatives to H that also entail E after becoming aware of the logical fact that H entails E. Just add an irrelevant conjunct to H or make a Goodman-style move. If one does so, then the probability of H goes down. Hence, a scientist who has coherent beliefs and who reasons in this way will hold the view that H is not confirmed.

[7]Note, though, that a' has to satisfy inequality (13).

Hence,

$$\overline{\alpha}\, a\, \overline{q} - \overline{\alpha'}\, a'\, \overline{q'} = 0\,. \tag{12}$$

Next, we conclude from eq. (12) that $\overline{\alpha}\, a\, \overline{q} = \overline{\alpha'}\, a'\, \overline{q'} < a'$. Hence, we obtain

$$a' > \overline{\alpha}\, a\, \overline{q}\,, \tag{13}$$

as our third and final constraint on P'. Inequality (13) tells us that the scientist cannot reduce a' to an arbitrarily low value. A rational agent should always take the possible existence of an alternative theory that entails E or the possibility that E resulted from a chance mechanism into account.
In this section we will only assume that the probability of E does not change after the scientist becomes aware of the logical fact that H entails E (and that other available theories do not entail E) and after she lowered the probability of A in turn. In the next section, we will focus on the specific case that $P(\mathrm{E}) = P'(\mathrm{E})$ is close to 1, which is the situation in the Problem of Old Evidence.
We can now show the following theorem (proof in the Appendix).

Theorem: *Consider the Bayesian Network in Figure 1 with the prior probability distribution P from eq. (6). We furthermore assume that (i) the posterior probability distribution P' is defined over the same Bayesian Network, (ii) the posterior distribution P' is constrained by eqs. (8), (12) and (13), and (iii) P' minimizes the Kullback-Leibler divergence between P' and P. Let $\Delta := P'(\mathrm{H}) - P(\mathrm{H})$. Then*

$$\Delta = \frac{q}{\alpha + \overline{\alpha}\, q} \cdot \overline{\alpha}\,(a - a') \tag{14}$$

Hence, $\Delta > 0$ if and only if $a' < a$. And so we conclude that H is confirmed in the considered situation although the probability of the evidence does not change in the course of deliberation. Note that H is not directly confirmed by E. The confirmation is *indirect* as we first become aware of the logical fact that H entails E (and that other available theories do not entail E) and then, in turn, reduce the probability of A. This, then, results in an increase in the probability of H if we determine the posterior probability distribution by minimizing the Kullback-Leibler divergence between the posterior probability distribution and the prior probability distribution.

3 The Problem of Old Evidence

Let us now turn to the Problem of Old Evidence. So far we have only assumed that the probability of E does not change in the course of deliberation. In the old-evidence situation, more specifically, the probability of E is close to 1. So let us set

$$P(\mathrm{E}) = 1 - \epsilon \tag{15}$$

with $\epsilon \ll 1$. Next, we observe from eq. (7) that

$$\epsilon = \overline{\alpha}\, a\, \overline{q}\,, \tag{16}$$

which has to be small. To achieve this, we set

$$q = 1 - \eta \tag{17}$$

with $\eta \ll 1$. Hence, $\epsilon = \overline{\alpha}\, \eta\, a < \eta \ll 1$. We do not have to impose any further constraints on the values of a and α. Note that the choice (17) makes a lot of sense: The scientist knows that $P(\mathrm{E})$ is close to 1, independently of whether H is true or false: It is simply a contingent fact that E obtains (note that ϵ and η can be set to an arbitrarily small value and so our proposed solution of the Problem of Old Evidence works, as we will see, for all practical purposes). However, the scientist does not know whether or not H is true, and so she should set $P(\mathrm{E}|\mathrm{H})$ and $P(\mathrm{E}|\neg\mathrm{H})$ to 1 or to a value very close to 1 (again, there is no difference between the two assignments for all practical purposes). It is important to note that setting $P(\mathrm{E}|\mathrm{H})$ and $P(\mathrm{E}|\neg\mathrm{H})$ to 1 (or to a value close to 1) does not mean that the scientist knows that E is a deductive consequence of H: In the present case, the scientist has to make these assignments simply in order to have coherent beliefs.[8]

To proceed, let us plot Δ as a function of α for plausible values of a, a' and q. Figure 2 shows that one gets a considerable amount of confirmation (for $\alpha = .2$, for example, the probability rises by .15). It is also plausible, as Figure 2 suggests, that one gets more confirmation for smaller values of α: if we strongly believe that there is an alternative to H that entails E or that E is the result of a chance mechanism and if we also strongly believe that H is false if there is an alternative to H or if E is the result of a chance mechanism, then we will be quite impressed by becoming aware of the logical fact that H entails E. Hence, *ceteris paribus*, smaller values of α lead to more confirmation of H than larger values of α.

All this can also be seen analytically. To do so, we expand Δ in a Taylor-series up to zeroth order in η and obtain:

$$\Delta = \overline{\alpha}\,(a - a') + \mathcal{O}(\eta) \tag{18}$$

[8] Note that $P(\mathrm{E}) = 1$ implies that $P(\mathrm{E}|\mathrm{H_i}) = 1$ for every element H_i of a partition H_1, H_2,... with a non-vanishing prior probability. However, from $P(\mathrm{E}) = 1 - \epsilon$ we cannot infer that $P(\mathrm{E}|\mathrm{H_i}) = 1 - \epsilon'$ with $\epsilon' \ll 1$. In the present case, however, the probability of H (as well as the probability of $\neg$H) is neither very small nor very large. And so the scientist has to set $P(\mathrm{E}|\mathrm{H})$ and $P(\mathrm{E}|\neg\mathrm{H})$ to 1 (or to a value close to 1) in order to have coherent beliefs.

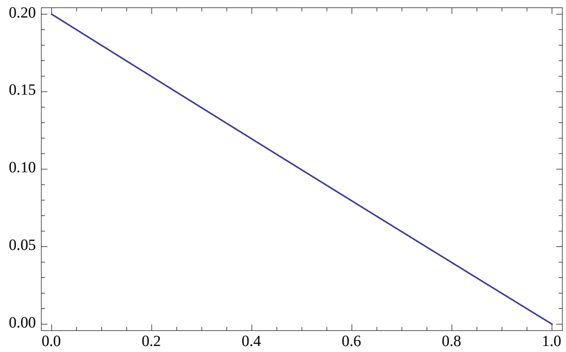

Figure 2: Δ as a function of α for $a = .8, a' = .6$ and $q = .99$.

Hence, H is more confirmed, the smaller the value of α (for fixed values of a and $a' > \overline{\alpha}\, \eta\, a$).
Note that we can take the limit $\eta \to 0$ from eq. (18) and obtain

$$\Delta = \overline{\alpha}\, (a - a'). \tag{19}$$

This result is approached as $P(\mathrm{E})$ approaches 1, so that our assumption that $P(\mathrm{E}) = 1 - \epsilon$ was only a mathematical trick that we had to make to proceed with the calculation. Interestingly, however, eq. (19) can also be obtained if one uses Jeffrey conditionalization in a straightforward way. To do so, we calculate the posterior probability of H after learning that the probability of A shifted:

$$\begin{aligned} P^*(\mathrm{H}) &= P(\mathrm{H}|\mathrm{A})\, P^*(\mathrm{A}) + P(\mathrm{H}|\neg\mathrm{A})\, P^*(\neg\mathrm{A}) \\ &= \alpha\, a' + \overline{a'}\,. \end{aligned} \tag{20}$$

With eq. (4), we obtain

$$\begin{aligned} \Delta^* &:= P^*(\mathrm{H}) - P(\mathrm{H}) \\ &= \overline{\alpha}\, (a - a')\,. \end{aligned} \tag{21}$$

Note that this calculation assumes that α is fixed and therefore does not change in the course of deliberation.[9] In hindsight, this assumption is justified as it turns out that α changes only slightly (details in the Appendix). It was, however, not at all clear from the beginning that this is the case, and so we were not justified to use Jeffrey conditionalization in the first place. Conditionalization and Jeffrey conditionalization often work, but they do not always work. The more general procedure to rationally change one's beliefs is to use the method of minimizing the Kullback-Leibler divergence taking into account the new information as a constraint on the posterior probability distribution.

[9] This can be seen by comparing eqs. (20) and (34).

In closing, let us note that our solution to the Problem of Old Evidence also shows that old evidence typically disconfirms the theory that preceded H. For example, the advanced perihelion of Mercury (= E) disconfirms Newtonian Mechanics (= H'). As $P(\mathrm{E}) = 1 - \epsilon$, we set (as before) $P(\mathrm{E}|\mathrm{H}')$ and $P(\mathrm{E}|\neg\mathrm{H}')$ to 1 or to a value close to 1. The scientist then becomes aware of the logical fact that H' does *not* entail E. She therefore increases the probability of A, i.e. she sets $a' > a$: It is now more probable that there is an alternative to H' that entails E or that E is the result of a chance mechanism. In fact, she has to set $a' = 1$ as she is now certain that there is either another theory that entails E or that E is the result of a chance mechanism. Hence, according to eq. (19), H' is disconfirmed and $P'(\mathrm{H}') < P(\mathrm{H}')$. We conjecture that it will be difficult to plausibly show this in Howson's counterfactual approach (Howson 1991).

4 Concluding Remark

Our argument crucially depends on the assumption that scientists believe the proposition A with a certain probability and that they change the corresponding probability once they become aware of the logical fact that H entails E (and that other available theories do not entail E). This is an empirical assumption, and it has be be investigated in detail whether it is true or false for concrete examples. We leave this for another occasion.

Acknowledgements

Thanks to Aidan Lyon for an eye-opening discussion. I would also like to thank Seamus Bradley, Erik Curiel, Igor Douven, Clark Glymour, Jim Joyce, Jan Sprenger, and Greg Wheeler for helpful feedback on an earlier draft.

A Appendix: Proof of the Theorem

Let P and P' be two probability distributions. The Kullback-Leibler divergence $D_{KL}(P'||P)$ between P' and P is defined as

$$D_{KL}(P'||P) := \sum_{i=1}^{n} P'(S_i) \log \frac{P'(S_i)}{P(S_i)} \, . \tag{22}$$

Here $S_1, \ldots, S_n$ be the possible values of a random variable S over which probability distributions P' and P are defined.

Using eqs. (6) and (10), we obtain:

$$
\begin{aligned}
D_{KL}(P'||P) &:= \sum_{A,H,E} P'(A,H,E) \log \frac{P'(A,H,E)}{P(A,H,E)} \\
&= \alpha'\,a' \log \frac{\alpha'\,a'}{\alpha\,a} + \overline{\alpha'}\,a'\,q' \log \frac{\overline{\alpha'}\,a'\,q'}{\overline{\alpha}\,a\,q} \qquad (23)\\
&\quad + \overline{\alpha'}\,a'\,\overline{q'} \log \frac{\overline{\alpha'}\,a'\,\overline{q'}}{\overline{\alpha}\,a\,\overline{q}} + \overline{a'} \log \frac{\overline{a'}}{\overline{a}} \\
&= a' \, \log \frac{a'}{a} + \overline{a'} \, \log \frac{\overline{a'}}{\overline{a}} + a' \left(\alpha' \, \log \frac{\alpha'}{\alpha} + \overline{\alpha'} \, \log \frac{\overline{\alpha'}}{\overline{\alpha}} \right) \qquad (24)\\
&\quad + \overline{\alpha'}\,a' \left(q' \, \log \frac{q'}{q} + \overline{q'} \, \log \frac{\overline{q'}}{\overline{q}} \right) \qquad (25)
\end{aligned}
$$

Next, we minimize

$$
L := D_{KL}(P'||P) + \lambda(\overline{\alpha}\,a\,\overline{q} - \overline{\alpha'}\,a'\,\overline{q'}) \tag{26}
$$

with respect to α' and q'. Here λ is a Lagrange multiplier, and the expression in the bracket takes the constraint from eq. (12) into account.
To find the minimum, we first differentiate L with respect to q' and obtain:

$$
\frac{\partial L}{\partial q'} = \overline{\alpha'}\,a' \left(\log \left(\frac{q'}{\overline{q'}} \, \frac{\overline{q}}{q} \right) + \lambda \right) \tag{27}
$$

Setting this expression equal to zero and noting that $a' > 0$ and $\alpha' < 1$, we obtain

$$
q' = \frac{q}{q + \overline{q}\,\mathrm{e}^{\lambda}} \,. \tag{28}
$$

With this, L simplifies to

$$
\begin{aligned}
L &= a' \, \log \frac{a'}{a} + \overline{a'} \, \log \frac{\overline{a'}}{\overline{a}} + a' \left(\alpha' \, \log \frac{\alpha'}{\alpha} + \overline{\alpha'} \, \log \frac{\overline{\alpha'}}{\overline{\alpha}} \right) \\
&\quad - \overline{\alpha'}\,a' \log(q + \overline{q}\,\mathrm{e}^{\lambda}) + \overline{\alpha}\,\lambda\,a\,\overline{q} \qquad (29)
\end{aligned}
$$

Next, we differentiate this expression with respect to α' and obtain

$$
\frac{\partial L}{\partial \alpha'} = a' \log \left(\frac{\alpha'}{\overline{\alpha'}} \, \frac{\overline{\alpha}}{\alpha} \, (q + \overline{q}\,\mathrm{e}^{\lambda}) \right) \,. \tag{30}
$$

Setting this expression equal to zero and noting that $a' > 0$, we obtain

$$
\alpha' = \frac{\alpha}{\alpha + \overline{\alpha}\,(q + \overline{q}\,\mathrm{e}^{\lambda})} \,. \tag{31}
$$

Inserting eqs. (28) and (31) into eq. (12), we obtain

$$\mathrm{e}^{\lambda} = \frac{(\alpha + \overline{\alpha}\, q)\, a}{a' - \overline{\alpha}\, a\, \overline{q}}\,. \tag{32}$$

Note that the denominator in eq. (32) is always greater than zero because of the constraint expressed in eq. (12).
Inserting eq. (32) into eqs. (28) and (31), after a short calculation we obtain

$$\alpha' = \frac{\alpha}{a'} \cdot \frac{a' - \overline{\alpha}\, a\, \overline{q}}{\alpha + \overline{\alpha}\, q} \quad , \quad q' = \frac{(a' - \overline{\alpha}\, a\, \overline{q})\, q}{a'\, q + \alpha\, a\, \overline{q}}\,. \tag{33}$$

This completes the calculation of the posterior probability distribution.
Let us now explore under which conditions the posterior probability of H, i.e. $P'(\mathrm{H})$ is greater than the prior probability of H, i.e. $P(\mathrm{H})$. That is, let us ask under which conditions H is confirmed. To do so, we calculate

$$P'(\mathrm{H}) = \alpha'\, a' + \overline{a'}\,. \tag{34}$$

Hence, using eq. (4), we finally obtain

$$\begin{aligned} \Delta &:= P'(\mathrm{H}) - P(\mathrm{H}) \\ &= \alpha'\, a' - \alpha\, a + (a - a') \\ &= \frac{q}{\alpha + \overline{\alpha}\, q} \cdot \overline{\alpha}\, (a - a')\,, \end{aligned} \tag{35}$$

which completes the proof of the theorem.

Let us finally consider α' and q' for the old evidence situation. We will show that $\alpha' \leq \alpha$ and $q' \leq q$ for $a > a'$. Setting $a' = \overline{\alpha}\, \eta\, a + \delta$ with $\delta > 0$ (see eq. (13)), we obtain

$$\begin{aligned} \frac{\alpha'}{\alpha} &= \frac{\delta}{(\overline{\alpha}\, \eta\, a + \delta)(\alpha + \overline{\alpha}\, \overline{\eta})} = \frac{\delta}{\delta + \overline{\alpha}\, \eta\, (a - \overline{\alpha}\, \eta\, a - \delta)} \\ &= \frac{\delta}{\delta + \overline{\alpha}\, \eta\, (a - a')} \leq 1\,. \end{aligned} \tag{36}$$

Similarly,

$$\begin{aligned} \frac{q'}{q} &= \frac{\delta}{(\overline{\alpha}\, \eta\, a + \delta)\, \overline{\eta} + \alpha\, \eta\, a} = \frac{\delta}{\delta + \eta\, (a - \overline{\alpha}\, \eta\, a - \delta)} \\ &= \frac{\delta}{\delta + \eta\, (a - a')} \leq 1\,. \end{aligned} \tag{37}$$

Hence, for $\eta \ll \delta$ and $a > a'$, we find that $\alpha' \lesssim \alpha$ and $q' \lesssim q$.

References

Bovens, L. and S. Hartmann (2003). *Bayesian Epistemology.* Oxford: Oxford University Press.

Diaconis, P. and S. Zabell (1982). Updating Subjective Probability, *Journal of the American Statistical Association* 77: 822–830.

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

Fitelson, B. (2004). Earman on Old Evidence and Measures of Confirmation. Unpublished manuscript. University of California-Berkeley.

Garber, D. (1983). Old Evidence and Logical Omniscience in Bayesian Confirmation Theory. In: J. Earman (ed.): *Testing Scientific Theories.* Minneapolis: The University of Minnesota Press, pp. 99–131.

Glymour, C. (1980). Why I am not a Bayesian. In: *Theory and Evidence.* Princeton: Princeton University Press, pp. 63–93.

Hartmann, S. and S. Rafiee-Rad (2014). Learning Conditionals. MCMP Working Paper.

Howson, C. (1991). The Old Evidence Problem. *The British Journal for the Philosophy of Science* 42: 547–555.

Norton, J. D. (2008). Ignorance and Indifference. *Philosophy of Science* 75: 45–68.

Norton, J. D. (2011). Challenges to Bayesian Confirmation Theory. In: P. Bandyopadhyay and M. Forster (eds.): *Handbook of the Philosophy of Science, Vol. 7: Philosophy of Statistics*. Amsterdam: Elsevier, pp. 391–440.