Bayesians sometimes cannot ignore even very implausible theories

(even ones that have not yet been thought of)

Branden Fitelson
Department of Philosophy
University of California–Berkeley
branden@fitelson.org

Neil Thomason
Department of Philosophy
University of Melbourne
neilt@unimelb.edu.au

June 7, 2007

ABSTRACT: In applying Bayes's theorem to the history of science, Bayesians sometimes assume – often without argument – that they can safely ignore very implausible theories. This assumption is false, both in that it can seriously distort the history of science as well as the mathematics and the applicability of Bayes's theorem. There are intuitively very plausible counter-examples. In fact, one can ignore very implausible or unknown theories only if at least one of two conditions is satisfied: (i) one is certain that there are no unknown theories which explain the phenomenon in question, or (ii) the likelihood of at least one of the known theories used in the calculation of the posterior is reasonably large. Often in the history of science, a very surprising phenomenon is observed, and neither of these criteria is satisfied.

1 Introduction

Bayes's Theorem comes in various flavors (see [7] for a nice survey). Presently, we will make use of the following (standard) version of Bayes's Theorem:

Bayes's Theorem. $\Pr(T \mid E) = \dfrac{\Pr(E \mid T) \cdot \Pr(T)}{\Pr(E \mid T) \cdot \Pr(T) + \Pr(E \mid \sim T) \cdot \Pr(\sim T)}$

In words, this version of Bayes's Theorem says that the conditional probability of a theory T, given an evidential proposition E (which we'll just call the

posterior of T, for short) is a function of four quantities: (i) the *likelihood* of T: $\Pr(E \mid T)$, (ii) the likelihood of the denial of T: $\Pr(E \mid \sim T)$, the unconditional probability of T (which we'll just call the *prior* of T, for short): $\Pr(T)$, and the prior of $\sim T$: $\Pr(\sim T)$. This paper can be read as a cautionary tale about the perils of trying to *approximate* the posterior of T, using Bayes's Theorem, together with *approximations of* some of the quantities (i)–(iv). Specifically, we will focus on the use of Bayes's Theorem, together with *approximations of* the quantities [(ii) and (iv)] involving the denial of T: $\Pr(\sim T)$, $\Pr(E \mid \sim T)$. Such approximations are often used in Bayesian philosophy (and history) of science.

It is well-known that Bayesians have a problem dealing with the likelihoods of unknown (unthought of) theories (see, for instance [3, Ch. 7]). If you have no idea of what a theory is, how can you say anything about the probability of a phenomenon (E) given that unknown theory? Bayesians sometimes get around this concern by assuming that sum of the prior probabilities of the known theories is (very) high, thus making the prior probabilities of any unknown theories (very) low. In applying Bayes's theorem to the history of science, Bayesians of this ilk sometimes tacitly rely — often without argument — on the following problematic assumption:

Highly-Implausible-Theories-Irrelevant (HITI) Assumption[1]:

> Given the prior probabilities and likelihoods of all known theories with respect to evidence E, and given that the sum of the prior probabilities of the known theories is close to 1, the posterior $\Pr(T \mid E)$ for any theory T can be approximated to a high degree of accuracy. I.e., for all practical purposes, under those conditions one can ignore very improbable (off the wall) alternatives to T, whether they be known or unknown.

While many mathematical examples *do* satisfy the HITI, other mathematical examples (and, more interestingly, *much of the history of science*) do *not*. The falsity of HITI adds additional complications for applying Bayes's theorem to the history of science and elsewhere.

[1] For some examples, of varying degrees of explicitness, see [1, p. 182], [2], [6, p. 137], [3, p. 84], and [10, *passim*]. Many of the confirmation-theoretic examples we know of come from discussions of Bayesian solutions to the Quine/Duhem Problem, but the issue obviously also arises elsewhere in Bayesian confirmation theory. Moreover, similar problems concerning "ignoring implausible alternatives" also arise in the more general context of Bayesian decision theory. See [8] for discussion.

2 A Truth Related to the HITI Assumption

Although the HITI Assumption is false (as we will demonstrate shortly), there is a related truth of some intrinsic interest that we will examine first. The related truth is the following:

> If one can accurately approximate the prior probability of the denial $\sim T$ of T (i.e., the disjunction of all the alternatives to T), then one can also accurately approximate the likelihood of $\sim T$ (relative to E), using one's approximation of the prior probability of $\sim T$.

Let $\Pr^*(\sim T)$ be an approximation of $\Pr(\sim T)$ calculated by adding the prior probabilities of the most probable theories inconsistent with T. And, let $\Pr^*(E \mid \sim T)$ be an approximation of the likelihood of $\sim T$ [i.e., the disjunction of all alternatives to T] relative to evidence E, obtained using $\Pr^*(\sim T)$ in place of $\Pr(\sim T)$. We will now show that, for any small difference δ we chose,

$$\textit{If } |\Pr(\sim T) - \Pr^*(\sim T)| < \delta \cdot \Pr(\sim T), \textit{ then } |\Pr(E \mid \sim T) - \Pr^*(E \mid \sim T)| < \delta.$$

Here is the proof. Assume there is a finite number of mutually exclusive theories T, T_1, T_2, …, T_j, …, T_n, where T_1, T_2, …, T_n are all the alternative theories to T.[2] The sum of the prior probabilities of the alternatives is the probability that T is false:

$$\Pr(T_1) + \Pr(T_2) + \cdots + \Pr(T_j) + \cdots + \Pr(T_n) = \Pr(\sim T)$$

For each theory, assume that we know its prior probability and its likelihood (relative to E). We can calculate $\Pr(\sim T)$ in two different ways, because

$$\Pr(\sim T) = \Pr(T_1) + \Pr(T_2) + \cdots + \Pr(T_n)$$

and

$$\Pr(\sim T) = 1 - \Pr(T)$$

We also know the likelihood of $\sim T$ (relative to E), since

$$\Pr(E|\sim T) = \frac{\Pr(E \mid T_1) \cdot \Pr(T_1) + \Pr(E \mid T_2) \cdot \Pr(T_2) + \cdots + \Pr(E \mid T_n) \cdot \Pr(T_n)}{\Pr(\sim T)}$$

Now, since we know $\Pr(T)$, $\Pr(E \mid T)$, $\Pr(\sim T)$ and $\Pr(E \mid \sim T)$, we have enough information to use Bayes's theorem to calculate $\Pr(T \mid E)$. Next, let

[2] We assume here that all alternatives to T are *incompatible* with T, and that there are only *finitely many* alternatives to T. These are standard Bayesian modeling assumptions. The result here could be generalized to various other sorts of cases, but this would unnecessarily complicate the present discussion.

us approximate $\Pr(\sim T)$ as closely as we desire, as $\Pr^*(\sim T)$. First, we order all the alternative theories to T in monotonic decreasing prior probability: $\Pr(T_1) \geq \Pr(T_2) \geq \cdots \geq \Pr(T_n)$. We want $\Pr^*(\sim T)$ to be within $\delta \cdot \Pr(\sim T)$ of $\Pr(\sim T)$; that is, $|\Pr(\sim T) - \Pr^*(\sim T)| < \delta \cdot \Pr(\sim T)$. Since we know $\Pr(\sim T)$ precisely, we can get the desired approximation by adding $\Pr(T_1) + \Pr(T_2) + \cdots$ until we reach a theory T_k such that:

$$\Pr(\sim T) - (\Pr(T_1) + \Pr(T_2) + \cdots + \Pr(T_k)) < \delta \cdot \Pr(\sim T).$$

Depending on the facts of the case and the desired degree of approximation, T_k may or may not be the same theory as T_n. First, note the triviality that given any theory, the probability of any data can never be greater than 1. That is, for any theory T_i, $\Pr(E, |T_i) \leq 1$. Therefore for any theory T_i,

$$\Pr(T_i) \cdot \Pr(E \mid T_i) \leq \Pr(T_i)$$

and so

$$\begin{aligned}
\Pr(E \mid T_{k+1}) \Pr(T_{k+1}) &\leq \Pr(T_{k+1}) \\
\Pr(E \mid T_{k+2}) \Pr(T_{k+2}) &\leq \Pr(T_{k+2}) \\
&\vdots \\
\Pr(E \mid T_n) \cdot \Pr(T_n) &\leq \Pr(T_n)
\end{aligned}$$

Therefore:

$$\Pr(E \mid T_{k+1}) \cdot \Pr(T_{k+1}) + \cdots + \Pr(E \mid T_n) \cdot \Pr(T_n) \leq \Pr(T_{k+1}) + \cdots + \Pr(T_n)$$

Since we have ordered our theories such that

$$\Pr(T_{k+1}) + \cdots + \Pr(T_n) < \delta \cdot \Pr(\sim T),$$

it follows that

$$\Pr(E \mid T_{k+1}) \cdot \Pr(T_{k+1}) + \cdots + \Pr(E \mid T_n) \cdot \Pr(T_n) < \delta \cdot \Pr(\sim T).$$

And, since

$$|\Pr(E \mid \sim T) - \Pr^*(E \mid \sim T)| = \frac{\Pr(E \mid T_{k+1}) \cdot \Pr(T_{k+1}) + \cdots + \Pr(E \mid T_n) \cdot \Pr(T_n)}{\Pr(\sim T)},$$

it follows that

$$|\Pr(E \mid \sim T) - \Pr^*(E \mid \sim T)| < \delta. \quad \square$$

Therefore, provided that we can approximate $\Pr(\sim T)$ to any degree of accuracy we want [by $\Pr^*(\sim T)$], we can also approximate $\Pr(E \mid \sim T)$ to any degree of accuracy we want [using $\Pr^*(\sim T)$].[3] Now we are in the happy position of knowing $\Pr(T)$, $\Pr(E \mid T)$ and $\Pr(\sim T)$, and we are able to approximate $\Pr(E \mid \sim T)$ as accurately as want. But all is not beer and skittles. Although these four values are all the independent variables required for using Bayes's Theorem to calculate $\Pr(T \mid E)$, that is no guarantee that $\Pr^*(T, |E)$ will be anywhere near $\Pr(T \mid E)$. First, let us examine two counter-examples to HITI and then turn to the general conditions under which highly implausible theories cannot be ignored.

3 Two Counter-examples to the HITI Assumption

3.1 A Big Urn Case

You have a very large urn in front of you. For reasons we need not go into, the content of this urn was generated by flipping a fair coin. If the coin landed heads on the first flip, the urn was filled as per "The A Theory" described below. If the coin landed heads on the second flip, the urn was filled as per "The B Theory." And so on through the 26^{th} flip. If the coin did not land heads on the first twenty-six flips, the process is started over until it lands heads. By the time you are involved, the process has been completed and you have in front of you an urn filled with exactly 10^{15} balls. You know the prior probability of each theory. Because it is possible, albeit extremely unlikely, that you will get more than 26 consecutive heads: the probability that the A Theory is correct is $0.5 + (0.5)^{27} + (0.5)^{54} + \cdots$, which is approximately 0.5. The probability that the B Theory is correct is $0.25 + (0.25)^{28} + (0.25)^{55} + \cdots$, which is approximately 0.25. In general, the probability that the i^{th} theory is correct is $(0.5) \cdot i + (0.5)^{i+26} + \cdots$, which is approximately $(0.5) \cdot i$. Finally, the probability that the Z theory is correct is $(0.5)^{26} + \cdots$, which is approximately $1.5 \cdot 10^{-8}$. Here are the theories:

> The A Theory: The urn was filled with $10^{15} - 1$ balls labeled "A" and one labeled "Z".
>
> The B Theory: The urn was filled with $10^{15} - 1$ balls labeled "B", and one labeled "Z".

[3] As a result, Bayesian confirmation theorists who measure degree of confirmation using the likelihood difference measure: $\Pr(E \mid T) - \Pr(E \mid \sim T)$ (e.g., [9, p. 252]) *can* safely ignore very implausible alternatives (in general). However, the likelihood difference is an inadequate measure of confirmation in salient contexts (see, *e.g.*, [4, fn. 26]).

> ⋮
>
> The Z Theory: The urn was filled with 10^{15} balls labeled "Z" and no ball with any other label.

Calculating the priors and posteriors is straightforward:

$$\Pr(A) \approx 0.5$$

$$\Pr(Z \mid A) = 10^{-15}$$

$$\Pr(B) \approx 0.25$$

$$\Pr(Z \mid B) = 10^{-15}$$

$$\Pr(C) \approx 0.125$$

$$\Pr(Z \mid C) = 10^{-15}$$

$$\vdots$$

$$p(Z) \approx 1.5 \cdot 10^{-8}$$

$$\Pr(Z \mid Z) = 1$$

$$\begin{aligned}\Pr(Z \mid \sim A) &= \Pr(Z \mid B) \cdot \Pr(B \mid \sim A) + \Pr(Z \mid C) \cdot \Pr(C \mid \sim A) + \cdots + \Pr(Z \mid Z) \cdot \Pr(Z \mid \sim A) \\ &\approx 10^{-15} \cdot 0.5 + 10^{-15} \cdot 0.25 + 10^{-15} \cdot 0.125 + \cdots + 1 \cdot 3 \cdot 10^{-8} \\ &\approx 3.0000001 \cdot 10^{-8}\end{aligned}$$

Suppose that you draw randomly from this urn and, to your considerable surprise, you get a "Z" ball. Intuitively, the probability of Theory A drops dramatically; Bayes's Theorem tells a similar story:

$$\begin{aligned}\Pr(A \mid Z) &= \frac{\Pr(Z \mid A) \cdot \Pr(A)}{\Pr(Z \mid A) \cdot \Pr(A) + \Pr(Z \mid \sim A) \cdot \Pr(\sim A)} \\ &\approx \frac{10^{-15} \cdot 0.5}{10^{-15} \cdot 0.5 + 3.0000001 \cdot 10^{-8} \cdot 0.5} \\ &\approx 3 \cdot 10^{-7}\end{aligned}$$

Now, we need only calculate $\Pr^*(A \mid Z)$. Suppose that you want your estimate $\Pr^*(\sim A)$ to be within one million parts of $\Pr(\sim A)$ in the above sense.[4] That is, you want $|\Pr(\sim A) - \Pr^*(\sim A)| < \delta \cdot \Pr(\sim A)$, where $\delta = 10^{-6}$.

[4] If you are inclined to think that one chance in a million is not a close enough approximation, you can easily modify our example by taking a leaf from Dr. Seuss. Add a finite

As it happens, the sum of the probabilities of theories B through W is easily within one million parts of $\Pr(\sim A)$ in this sense. That is,

$$\Pr(\sim A) - [\Pr(B) + \Pr(C) + \Pr(D) + \cdots + \Pr(W)] < \delta \cdot \Pr(\sim A) = 5 \cdot 10^{-7}$$

Thus, we can safely define $\Pr^*(\sim A)$ in this way:

$$\Pr^*(\sim A) =_{\text{df}} \Pr(B) + \Pr(C) + \Pr(D) + \cdots + \Pr(W)$$

As a result,

$$\begin{aligned}\Pr^*(Z \mid \sim A) &= \Pr(Z \mid B) \cdot \Pr(B \mid \sim A) + \Pr(Z \mid C) \cdot \Pr(C \mid \sim A) + \cdots + \Pr(Z \mid W) \cdot \Pr(W, \sim A) \\ &\approx 10^{-15} \cdot 0.5 + 10^{-15} \cdot 0.25 + 10^{-15} \cdot 0.125 + \cdots + 10^{-15} \cdot 10^{-7} \\ &\approx 10^{-15}\end{aligned}$$

This gives us:

$$|\Pr(Z \mid \sim A) - \Pr^*(Z \mid \sim A)| \approx |10^{-7} - 10^{-15}| < \delta$$

Now, let us use Bayes's theorem to calculate $\Pr^*(A \mid Z)$:

$$\begin{aligned}\Pr^*(A \mid Z) &= \frac{\Pr(Z \mid A) \cdot \Pr(A)}{\Pr(Z \mid A) \cdot \Pr(A) + \Pr^*(Z \mid \sim A) \cdot \Pr^*(\sim A)} \\ &\approx \frac{10^{-15} \cdot 0.5}{10^{-15} \cdot 0.5 + 10^{-15} \cdot 0.5} \\ &\approx 0.5\end{aligned}$$

The approximation $\Pr^*(A \mid Z) \approx 0.5$ is not at all close to $\Pr(A \mid Z) \approx 3 \cdot 10^{-7}$. In fact, it is off by about 7 orders of magnitude. Further, although drawing the "Z" ball caused the actual probability of Theory A to collapse, it did not change the approximation of the probability of Theory A at all.

It is clear why $\Pr^*(A \mid Z)$ went so awry. After a "Z" ball is drawn, it is almost certain that the Z Theory is true and therefore that the A theory false. However, since the prior probability of the Z Theory was so low it

number of letters between "A" and "Z" — this technique is known as In-Before-Zebra. Creativity is required here. With each additional letter, for each theory increase the number of balls in the urn by one order of magnitude, all eponymously labeled. For example, if you add 10 new letters to the alphabet, increase the size of the hypothesized "R" urn to $10^{25} - 1$ "R" balls and one "Z" ball. With this process, one can drive δ down below any finite number you would feel safe ignoring. Of course, each new letter drives the probability of drawing a "Z" ball from the urn still further down. But, if you did draw a "Z" ball, the probability that Theory Z is true would be even closer to one.

was not included in the approximation $\Pr^*(\sim A)$, $\Pr(Z)$ played no role in the calculation of $\Pr^*(Z \mid \sim A)$ and thus of $\Pr^*(A \mid Z)$.

But, although the absolute values of $\Pr(Z \mid \sim A)$ and $\Pr^*(Z \mid \sim A)$ are very similar (they differ by less than δ, i.e., by less than 10^{-6}), their ratio is very large. In fact, since $\Pr(Z \mid \sim A) \approx 3 \cdot 10^{-7}$ and $\Pr^*(Z \mid \sim A) = 10^{-15}$, the likelihood ratio $\Pr^*(Z \mid \sim A) / \Pr(Z \mid \sim A)$ is about $3 \cdot 10^7$. This means that the genuine likelihood ratio $\Pr(Z \mid A) / \Pr(Z \mid \sim A)$ is about $3 \cdot 10^7$ times as large as the ersatz likelihood ratio $\Pr(Z \mid A) / \Pr^*(Z \mid \sim A)$. Because of this large discrepancy in likelihood ratios, $\Pr(A \mid E)$ when calculated against $\Pr^*(E \mid \sim A)$ will be very different than when calculated against $\Pr(E \mid \sim A)$. Since on most[5] Bayesian theories of confirmation, such large posterior probability (and likelihood ratio) discrepancies will lead to large discrepancies in judgments of degree of confirmation, most Bayesian confirmation theorists should eschew HITI.

Such phenomena do not arise only with urns. In fact, their *possibility* permeates science, even very well established theories.

3.2 The Spherical Earth Case

Let $T_{\text{sphere}} =_{\text{df}}$ The Earth is approximately a sphere and it does not rest on anything. And, assume (arguendo) that $\Pr(T_{\text{sphere}}) = 0.999999$.

There are a huge number of alternative theories to T_{sphere}, some already proposed, many hitherto unproposed. Let $\sim T_{\text{sphere}} =_{\text{df}}$ the set of all theories that hold that the Earth is not approximately spherical and/or does rest on something. Given our rather arbitrary estimation of $\Pr(T_{\text{sphere}})$ as $1 - 10^{-6}$, this roughly sets $\Pr(\sim T_{\text{sphere}})$ at about 10^{-6}. But nothing depends on the exact figure. If you'd like to make it even lower, that would be fine with us.

The set T_{sphere} can be divided into many very implausible theories:

> $T_{\text{sphere\&turtle}} =_{\text{df}}$ The Earth is approximately a sphere and it rests on the back of a turtle;
>
> $T_{\text{sphere\&2turtles}} =_{\text{df}}$ The Earth is approximately a sphere and it rests on the back of two turtles;
>
> $T_{\text{sphere\&2+turtles}} =_{\text{df}}$ The Earth is approximately a sphere and it rests on the back of more than two turtles. [This last theory is to prevent an

[5] See previous discussion of Nozick (fn. 3, above). See, also, [5], which discusses some related perils of using "approximations" in Bayesian confirmation theory.

infinite sequence of turtle theories. Such infinite sequences give rise to other problems that we would prefer not to deal with here.]

⋮

$T_{\text{sphere\&elephant}} =_{\text{df}}$ The Earth is approximately a sphere and it rests on the back of one or more elephants that rest on nothing.

⋮

$T_{\text{disk}} =_{\text{df}}$ The Earth is approximately a disk and it does not rest on anything;

$T_{\text{disk\&turtle}} =_{\text{df}}$ The Earth is approximately a disk and it rests on the back of a turtle;

$T_{\text{disk\&elephant\&turtle}} =_{\text{df}}$ The Earth is approximately a disk and it rests on the back of one or more elephants that rest on the back of a turtle.[6]

⋮

$T_{\text{torus}} =_{\text{df}}$ The Earth is approximately a torus and it does not rest on anything;

⋮

$T_{\text{tetrahedron}} =_{\text{df}}$ The Earth is approximately a tetrahedron it does not rest on anything;

⋮

$T_{\text{Möbius}} =_{\text{df}}$ The Earth is approximately a Möbius Strip and it does not rest on anything;

To ensure there are only a finite number of theories, we add a catchall theory:

$T_{\text{residual}} =_{\text{df}}$ All the above theories are false.

Even the most plausible of the many many alternatives to the Earth-is-a-Sphere-Resting-on-Nothing Theory is very improbable.

Now suppose that, every since the first satellites were launched, all astronauts, all photos, and all other measurements from space strongly indicate

[6]This theory has been advanced by Terry Pratchett in several recent publications.

that the Earth appears to have four corners, connected by roughly straight lines, ie, the Earth looks approximately like a tetrahedron — a theory so implausible that no one had bothered mentioning its possibility even to immediately dismiss it. It does not appear to rest on anything — not on turtles, not on a huge double-helix, not on giant reproductions of the Mona Lisa, ... Let us call this large amount of remarkable evidence E.

We don't know how to precisely determine what $\Pr(T_{\text{tetrahedron}})$ and $\Pr(E \mid T_{\text{tetrahedron}})$ are, but it seems obvious that the first is very small and the second very close to 1. So, for our purposes, let's just say that $\Pr(T_{\text{tetrahedron}}) = 10^{-7}$ and $\Pr(E \mid T_{\text{tetrahedron}}) = 1 - 10^{-12}$. $\Pr(E \mid T_{\text{tetrahedron}})$ isn't exactly 1 because it is always possible that light in certain bizarre circumstances reflects from a tetrahedron in such a way as to produces, e.g., the appearance of a sphere.

Our results don't require that these numbers are precise. All that we need are three very reasonable assumptions: (a) $\Pr(T_{\text{tetrahedron}})$ is so small that it is not included in calculating $\Pr^*(\sim T_{\text{sphere}})$, (b) $\Pr(E \mid T_{\text{tetrahedron}})$ is extremely close to 1, and (c) $\Pr(E \mid T_{\text{any non-tetrahedron theory}})$ is much, much smaller than $\Pr(E \mid T_{\text{tetrahedron}})$. We will spare you the calculations.

$$\Pr(T_{\text{sphere}} \mid E) \approx 10^{-3}$$

$$\Pr^*(T_{\text{sphere}} \mid E) \approx 0.99999$$

There is no doubt about it; $\Pr^*(T_{\text{sphere}} \mid E)$ is again a remarkably bad approximation of $\Pr(T_{\text{sphere}} \mid E)$. Such cases are easy to create once one is in the right mind-set — a stiff drink in front of the fireplace helps.

4 When can highly implausible data make $\Pr^*(T \mid E)$ a very poor approximation of $\Pr(T \mid E)$?

When can a theory excluded from calculating $\Pr^*(\sim T)$ result in $\Pr^*(T \mid E)$ being so different from $\Pr(T \mid E)$? As the examples indicate, $\Pr(T)$ being near 1 does not make T safe. The rough answer is that T must entail that E is very unlikely whereas there must be at least one very unlikely theory that says that E is very likely. The very unlikely theory must be so unlikely that it is not included in calculating $\Pr^*(\sim T)$.

To state the problem a bit more formally, we are looking for the general conditions for cases that satisfy these criteria, where it assumed that we know all priors and likelihoods:

(1) $|\Pr(\sim T) - \Pr^*(\sim T)| < \delta \cdot \Pr(\sim T)$

(2) $|\Pr(E \mid {\sim}T) - \Pr^*(E \mid {\sim}T)| < \delta$ (As we saw above, this follows from 1)

(3) $|\Pr(T \mid E) - \Pr^*(T \mid E)| \gg \delta$

and/or

(3′) $\dfrac{\Pr^*(T \mid E)}{\Pr(T \mid E)} \gg 1$

The choice of criterion between (3) and (3′) depends on whether one wants the absolute difference between the approximation and the actual value to be large or whether one wants their ratio to be high. Sometimes both (3) and (3′) are satisfied. First we need some terminology. Let us call all the highly implausible theories not included in calculating $\Pr^*({\sim}T)$ the "Residual" T_{Resid}. We know by definition that $\Pr(T_{\text{Resid}}) < \delta$ and the theorem above gives $\Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}}) < \delta$. Since all other factors are held constant to satisfy criteria (3) or (3′), $\Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}})$ must make a substantial difference to the result of Bayes's theorem. The relevant forms of Bayes's theorem will be:

$$\Pr(T \mid E) = \frac{\Pr(E \mid T) \cdot \Pr(T)}{\Pr(E \mid T) \cdot \Pr(T) + \Pr(E \mid {\sim}T) \cdot \Pr({\sim}T)}$$

$$= \frac{\Pr(E \mid T) \cdot \Pr(T)}{\Pr(E \mid T) \cdot \Pr(T) + \Pr^*(E \mid {\sim}T) \cdot \Pr^*({\sim}T) + \Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}})}$$

$$\Pr^*(T \mid E) = \frac{\Pr(E \mid T) \cdot \Pr(T)}{\Pr(E \mid T) \cdot \Pr(T) + \Pr^*(E \mid {\sim}T) \cdot \Pr^*({\sim}T)}$$

So, we are looking for the conditions under which adding $\Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}})$ to the denominator results in $\Pr(T \mid E)$ being very different from $\Pr^*(T \mid E)$. Since both $\Pr(T_{\text{Resid}})$ and $\Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}})$ are less than the very small δ, we need the conditions where adding a very small number to the denominator makes such a major difference to $\Pr(T \mid E)$.

This can happen iff

$$\frac{\Pr(E \mid {\sim}T_{\text{Resid}}) \cdot \Pr({\sim}T_{\text{Resid}})}{\Pr(E \mid T) \cdot \Pr(T) + \Pr^*(E \mid {\sim}T) \cdot \Pr^*({\sim}T)} \gg 1$$

That is, iff the rest of the denominator is much smaller than Residual. Of course, if the rest of the denominator is much smaller than the Residual, the numerator will be too. If $\Pr(E \mid T) \cdot \Pr(T) + \Pr^*(E \mid {\sim}T) \cdot \Pr^*({\sim}T)$ is largish relative to $\Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}})$, then adding $\Pr(E \mid T_{\text{Resid}}) \cdot \Pr(T_{\text{Resid}})$ to

the denominator will only slightly change the value of the denominator and thus of the overall fraction.

We can informally summarize these last few reflections, as follows.

One can legitimately ignore highly improbable theories in calculating the approximation $\Pr^*(T \mid E)$ iff at least one of these criteria is satisfied:

(i) For *at least one* of the theories T_i *used* in calculating $\Pr^*(T \mid E)$, the likelihood $\Pr(E \mid T_i)$ is reasonably *high*.

or

(ii) For *all* of the theories T_j *not* used in calculating $\Pr^*(T \mid E)$, the likelihood $\Pr(E \mid T_j)$ is reasonably low.

Of course, what "reasonably" means will somewhat depend on the purposes behind the calculations. The urn and sphere cases show how seriously things can go awry if neither of these criteria is satisfied. It is important to note that both criteria deal with likelihoods, which are generally taken to be the most objective and robust ingredients of calculations using Bayes's Theorem. As such, both subjective and objective Bayesians should see these criteria as meaningful and probative.

5 Does the falsity of HITI matter?

In one straightforward sense, the falsity of the HITI Assumption does not matter to the practice of science. When situations that satisfy the two criteria arise, many scientists see that they must take the initially very implausible theory very seriously.

But, its falsity does matter to Bayesian applications to science, particularly the history of science. No aspect of an adequate philosophy of science should rely on a fallacious mathematical inference. This truism applies *a fortiori* to approaches such as Bayesianism that get much of their strength from relying on well-grounded probability theory. Its arguments should not depend on a demonstrably false assumption. In its presently used form, HITI is demonstrably invalid. Bayesians can only use a modified HITI, which applies only when at least one of the following two criteria is satisfied:

(i) It is certain that there are no unknown theories according to which the phenomenon in question is very probable.

or

(ii) The probability of the phenomenon is reasonably large, given at least one of the known theories used in the calculation of the posterior.

We suspect that often in the history of science, neither criterion is satisfied.

References

[1] Dorling, Jon. "Bayesian Personalism, the Methodology of Research Programmes, and Duhem's Problem", *Studies in History and Philosophy of Science*, Vol. 10 (1979), pp. 177–187.

[2] Dorling, Jon. "Further Illustrations of the Bayesian Solution of DuhemÕs Problem". Unpublished manuscript, 1982.

[3] Earman, John. *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*, MIT Press, Cambridge, 1992.

[4] Fitelson, Branden. "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity", *Philosophy of Science* 66 (1999), S362–S378.

[5] Fitelson, Branden and Waterman, Andrew. "Comparative Bayesian Confirmation and the Quine–Duhem Problem: A Rejoinder to Strevens", *British Journal for the Philosophy of Science*, forthcoming, 2007.

[6] Howson, Colin and Urbach, Peter. *Scientific reasoning: the Bayesian approach* (2nd ed.) Open Court Press, Chicago, 1993.

[7] Joyce, James. "Bayes's Theorem" in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2004 Edition), URL = `http://plato.stanford.edu/archives/sum2004/entries/bayes-theorem/`

[8] Lance, Mark. "Subjective Probability and Acceptance", *Philosophical Studies*, Vol. 77 (1995), pp. 1147–179.

[9] Nozick, Robert. *Philosophical Explanations*, Harvard University Press, 1981.

[10] Strevens, Michael. "The Bayesian Treatment of Auxiliary Hypotheses", *British Journal for the Philosophy of Science*, Vol. 52 (2001), 515–537, pp. 1147—179.