Comparative Bayesian Confirmation and the Quine–Duhem Problem: A Rejoinder to Strevens

Branden Fitelson and Andrew Waterman

To appear in *The British Journal for the Philosophy of Science*

Abstract. By and large, we think Strevens's [6] is a useful reply to our original critique [2] of his paper on the Quine–Duhem (QD) problem [5]. But, we remain unsatisfied with several aspects of his reply (and his original paper). Ultimately, we do not think he properly addresses our most important worries. In this brief rejoinder, we explain our remaining worries, and we issue a revised challenge for Strevens's approach to QD.

1 Strevens's "Clarifications"

In section two of his reply, Strevens identifies several ways in which our critique may "mislead the reader". We accept some of these points as potentially misleading, but we would like to remark on three of the other "clarifications" he offers in this section.

First, Strevens points out that he does not assume or need to assume that e and $\neg(ha)$ are "equivalent" in any strong sense, but only that they are probabilistically indistinguishable in a local sense. This is true, but none of our arguments in [2] rely on any stronger notion of equivalence either, so there is no real issue here. Both parties can get by with mere local probabilistic indistinguishability. And, since we don't say precisely what kind of equivalence we are assuming, the charitable reading is the weakest one that makes the bulk of our claims true (and local probabilistic equivalance is certainly enough for *that*).

Second, Strevens laments that "by restricting their attention to the question of whether h or a is relatively more confirmed or disconfirmed by e, they ignore the most interesting claims in the paper, such as claim (2) from the previous section." We don't know what "most interesting" means. It is true that we chose to focus not on (2), but instead on the relative confirmation e provides for h *vs* a in QD problems. But, as we explained in our original critique, we think that the original QD problem is *about* the relative confirmational impact of e on h *vs* a. So, we don't think it is inappropriate to focus on this when critiquing a paper on QD. Indeed, most Bayesians seem to be in agreement with us on this score. Howson and Urbach [3, page 136, our notation] setup the QD problem in the following (canonical) way:

> Suppose a theory, h, and an auxiliary hypothesis, a, together imply an empirical consequence which is shown to be false by the observation of the outcome e. Let us assume that while the combination of ha is refuted by e, the two components taken individually are not refuted. We wish to consider the separate effects wrought on the probabilities of h and a by the adverse evidence e. The comparisons of interest are between $P(h \mid e)$ and $P(h)$, and ... $P(a \mid e)$ and $P(a)$.

The last sentence of this quote is crucial. It indicates that what matters are the *relative degrees of* ***relevance*** *confirmation conferred on* h vs a *by* e, and *not merely the posterior probabilities* of h *vs* a on e. So, while (2) might be "more interesting" *to Strevens*, it's unclear how this is supposed to translate into a response to *our* objections, which are motivated by the canonical *comparative confirmation-theoretic* formulation of QD.

Third, Strevens says he doesn't endorse any particular way of measuring confirmation (probabilsitically). This is fine, but failing to endorse a measure of confirmation does not constitute a response to the problems raised by our Theorem 3. In the next section, we will discuss the relationships between posterior probabilities, Bayesian (relevance) confirmation measures, and comparative Bayesian (relevance) confirmation in the context of the QD problem. From our point of view, this is the central set of issues raised by our critique, and we think they haven't been properly addressed (or addressed at all) in Strevens's reply.

2 Strevens's New-and-Improved "Negligibility Arguments"

The main crux of Strevens's reply has to do with (what he now calls) "the negligibility argument." Here, he clarifies the argument, which is very useful. He also argues that we misconstrue the argument, and that the argument does indeed go through as he originally desired, despite what we say in [2]. We think he is partly right (but partly misleading) on that score. In this section, we will get to the bottom of this central part of Strevens's reply.

Strevens's clarification of "the negligibility argument" is helpful, as it reveals that there are really the following *two distinct senses* of "approximation" being used in the argument.[1]

$x \approx_1 y$ iff $\frac{x}{y} = 1 + \epsilon$, for small ϵ.

$x \approx_2 y$ iff $|x - y| \leq \epsilon$, for small ϵ.

With these two distinct notions of approximation $\approx_1$ and $\approx_2$ in mind, Strevens clarifies the deductive special case of the negligibility argument as the following entailment:

$$(\dagger) \qquad \text{If } P(ha \mid e) = 0\text{, then } P(e \mid h\neg a) \approx_1 P(e \mid \neg(ha)) \Longrightarrow P(h \mid e) \approx_2 P(h \mid \neg(ha)).$$

In this part of the discussion, Strevens should make it clearer that $P(ha \mid e) = 0$ is a precondition of (†). Anyhow, this does become clearer when Strevens proves the following more general result, which subsumes the result (†) above:

$$(\ddagger) \qquad P(e \mid h\neg a) \approx_1 P(e \mid \neg(ha)) \Longrightarrow P(h \mid e) \approx_2 P(h \mid \neg(ha)) \cdot P(\neg(ha) \mid e) + P(ha \mid e).$$

While Strevens is correct that (†) and (‡) are clarified (and true!) renditions of the negligibility argument (and we thank him for correcting us on that score), his claim that they do "not assume any particular measure of degree of confirmation" is misleading. We would never claim that *these* kind of results are sensitive to choice of confirmation measure. What

[1]We should have been more careful in our original reconstruction of this part of his argument, and we now concede that our original Theorem 2 was not probative. But, in our defense, Strevens does admit that he was not crystal clear about the role of these two kinds of "approximation" in his original paper.

we are saying is that the *relevance* of results like (†) and (‡) to the QD problem needs further argument. If, as we (and many others) assume, an important aspect of resolutions of QD involves demarcating conditions under which *comparative* confirmation claims like $\mathfrak{c}(h,e) \geq \mathfrak{c}(a,e)$ come out true, then it is unclear why (‡) and (†) are *salient*. If one could show that (salient) measures of *degree of confirmation* were subject to clarified "negligibility theorems" akin to (†) and (‡), then *that* would be a response to our Theorem 3, which aimed to show that *even if posterior probabilties* $P(h \mid e)$ are subject to "negligibility theorems", it does not follow that *Bayesian (relevance) confirmation measures* $\mathfrak{c}(h,e)$ are. As we explain below, this *cannot* be shown, because it is *false*. Before we get to that, we first need to fix our old Theorem 3, in light of Strevens's clarification of the negligibility argument.

Unfortunately, our old Theorem 3 was operating under the false assumption that there was only one kind of approximation ($\approx_2$) being used in Strevens's arguments. So, to maintain our objections arising from Theorem 3, we need a new-and-improved Theorem 3, which is properly analogous to Strevens's new results (†) and (‡). Happily, such new-and-improved Theorems can be proven. We begin by discussing the deductive special case (all of our arguments concerning this special case can be lifted to the fully general case — see *fn.* 2). An analogous result to (†), but for Bayesian (relevance) confirmation measures, would be:

($\dagger_{\mathfrak{c}}$) If $P(ha \mid e) = 0$, then $P(e \mid h\neg a) \approx_1 P(e \mid \neg(ha)) \Longrightarrow \mathfrak{c}(h,e) \approx_2 \mathfrak{c}(h,\neg(ha))$.

Now, *if* ($\dagger_{\mathfrak{c}}$) were true for all salient Bayesian confirmation measures $\mathfrak{c}$, *then* our worries about how Strevens's results bear on the *comparative confirmational* QD question would be (largely) otiose. *But*, unfortunately, ($\dagger_{\mathfrak{c}}$) is *false* for many plausible measures of confirmation, including the likelihood-ratio measure $l(h,e) = P(e \mid h)/P(e \mid \neg h)$. That is, we have the following new-and-improved rendition of Theorem 3, along the lines of Strevens's clarified negligibility theorems (†) and (‡) (in the interest of brevity, we omit all proofs).

Theorem 3*. *Even if* $P(ha \mid e) = 0$ *and* $P(e \mid h\neg a) \approx_1 P(e \mid \neg(ha))$, it does *not* follow that $l(h,e) \approx_2 l(h,\neg(ha))$.

Indeed, one can provide an algorithm (similar to the one we used to establish our original Theorem 3) that will generate probability models such that all three of the following obtain:

(1) $P(ha \mid e) = 0$,

(2) $\dfrac{P(e \mid h\neg a)}{P(e \mid \neg(ha))} = 1 + \epsilon$, for ϵ as small as you like, and

(3) $|l(h,e) - l(h,\neg(ha))|$ is arbitrarily large (say, greater than $\frac{1}{\epsilon}$).[2]

Moreover, as one of us has recently argued elsewhere (see [1]), the likleihood-ratio measure l is *clearly superior* to the posterior probability (and many other existing candidate

[2] Our new Theorem 3* can be generalized to a (‡)–like result, as follows. Since the likelihood-ratio is just a function of the posterior (on e) and prior of h, we can compute $l'(h,e)$ — the "approximate" likelihood-ratio — as a function of the "approximate" posterior (which Strevens calls Q) and prior of h. Then, we can prove a theorem just like Theorem 3* above, but with $l(h,\neg(ha))$ replaced by $l'(h,e)$. It is also important to note that *no analogue of (†) or (‡) can be proven for* l. That is to say, *no matter what combination of* $\approx_1$ *and* $\approx_2$ *are used*, we cannot generate analogues of (†) or (‡) for l. Again, in order to save space, we omit all proofs.

measures) *in the context of explicating comparative* Bayesian-confirmation theoretic claims of the form $\mathfrak{c}(h,e) \geq \mathfrak{c}(a,e)$. That is, it is shown in [1] that $P(h \mid e) \geq P(a \mid e)$ is an *improper* explication of *comparative confirmation* claims of the form ⌜e favors h over a⌝, whereas $l(h,e) \geq l(a,e)$ is perfectly adequate for this purpose.[3] Strevens says that, unlike us, he "does not consider the selection of a single correct measure of confirmational relevance essential for work in confirmation theory." Contrary to what Strevens suggests here, *we* don't think that this is essential *in general* either. But, *in contexts such as these* — where we aim to establish *comparative* confirmation claims — we *do* think it is quite clear that $l(h,e)$ is far superior to $P(h \mid e)$ (and many other existing candidates) as a measure of confirmation. So, even in light of his helpful clarifications and replies, Strevens still owes us a response to our Theorem $3^{\star}$, which (in light of the arguments in [4] and [1]) seems to show that his results cannot undergird a proper, *comparative* Bayesian confirmation-theoretic resolution of the QD problem.

References

[1] B. Fitelson, 2007, "Likelihoodism, Bayesianism, and Comparative Confirmation," *Synthese*, forthcoming. URL: `http://fitelson.org/synthese.pdf`.

[2] B. Fitelson and A. Waterman, 2005, "Bayesianism and Auxiliary Hypotheses Revisited: A Reply to Strevens," *British Journal for the Philosophy of Science* **56**: 293–302.

[3] C. Howson and P. Urbach, 1993, *Scientific Reasoning: The Bayesian Approach* (Second Edition), Chicago: Open Court.

[4] K. Popper, 1954, "Degree of Confirmation," *The British Journal for the Philosophy of Science*, **5**: 143–149.

[5] M. Strevens, 2001, "The Bayesian Treatment of Auxiliary Hypotheses," *British Journal for the Philosophy of Science* **52**: 515–538.

[6] M. Strevens, 2005, "The Bayesian Treatment of Auxiliary Hypotheses: A Reply to Fitelson and Waterman", *British Journal for the Philosophy of Science* **56**: 913–918.

[3] We think it has been definitively established that a theory of favoring based on posterior probability comparisons is hopeless. A compelling argument to this effect was made over 50 years ago in this very journal by Karl Popper [4]. Moreover, we also think it is clear that a proper Bayesian theory of favoring can be formulated using likelihood-ratio comparisons. See [1] for various positive arguments in favor of l as a *comparatively adequate* Bayesian measure of confirmation (in contrast to other candidate measures).