Bayesian Confirmation and Auxiliary Hypotheses Revisited: A Reply to Strevens

Branden Fitelson and Andrew Waterman

Abstract. Michael Strevens [2001] has proposed an interesting and novel Bayesian analysis of the Quine-Duhem (Q–D) problem (*i.e.*, the problem of auxiliary hypotheses). Strevens's analysis involves the use of a simplifying idealization concerning the original Q–D problem. We will show that this idealization is far stronger than it might appear. Indeed, we argue that Strevens's idealization *over*simplifies the Q–D problem, and we propose a diagnosis of the source(s) of the oversimplification.

1 Some Background on Quine–Duhem

The Quine–Duhem (Q–D) problem (*i.e.*, the problem of auxiliary hypotheses) is traditionally seen as a problem for naive Hypothetico-Deductive (H-D) accounts of confirmation, according to which a piece of evidence O confirms a hypothesis H (roughly) just in case H entails O. Duhem was perhaps the first to clearly point out that no hypothesis *by itself* entails any interesting observation. Only hypotheses *conjoined with auxiliary hypotheses/assumptions* can entail interesting observations. So, in general, what we have is that $T = H \,\&\, A$ entails O, while H alone does not. Thus, if E is observed and E contradicts O, then the most we can say is that E refutes T. We cannot say which of H or A (or any other "parts" of T) is *individually* disconfirmed by E. As John Earman [1992, page 95] asks:

> ...if O is found to hold, to which parts of T can the praise for the successful prediction be attributed? If $\neg O$ is found to hold, on which parts of T can the blame for the unsuccessful prediction be laid?

Michael Strevens [2001] offers a Bayesian analysis of Q–D which is intended to provide a way to apportion blame between H and A, when E is observed.[1] We will interpret

[1]Most discussions of Q–D (Bayesian or otherwise) restrict their attention to cases of *deductive evidence* (in which H & A *entails* O). Strevens discusses both the traditional deductive Q–D problem, and the more general case in which E (merely) disconfirms, but does not refute, T. For an alternative non-deductive (and non-Bayesian) probabilistic treatment of Q–D, see Sober [2004]. For other Bayesian approaches that operate under the traditional presupposition of deductive evidence, see Dorling [1979] and Howson and Urbach [1993]. We will restrict our attention to problems we see in Strevens's treatment of the original (deductive) Q–D problem. We think that similar problems will arise for Strevens in the more general setting, but that is beyond the scope of this paper.

this as providing a way to (at least, qualitatively) compare $\mathfrak{c}(H, E)$ and $\mathfrak{c}(A, E)$, where $\mathfrak{c}(x, y)$ is a Bayesian measure of the degree to which y confirms (or supports) x. That is to say, we will examine the consequences of Strevens's approach with respect to the conditions under which $\mathfrak{c}(H, E) \geq \mathfrak{c}(A, E)$ in cases where H & A *entails* $\neg E$.

2 Strevens's Simplifying Idealization

The central element of Strevens's proposal is to simplify the original Q–D problem, by making use of the following idealization:

$$(\mathcal{I}) \qquad E \text{ is } \textit{equivalent to } \neg(H \,\&\, A)$$

$(\mathcal{I})$ asserts, not just that E entails that H and A aren't both true (as in the original Q–D problem), but also that E says *no more than this*. Strevens (p. 524) urges us to focus on the disconfirmation that E provides regarding H (*vs* A) "in virtue of *and only in virtue of* falsifying" $H \& A$. He goes on to argue that $(\mathcal{I})$ is not so strong as to *over*simplify the problem from the perspective of Bayesian confirmation theory. We disagree. Below, we (*i*) describe some consequences of $(\mathcal{I})$ that we think cast doubt on the legitimacy of Strevens's approach to Q–D, and (*ii*) offer a diagnosis of a possible source of the oversimplification implicit in Strevens's approach.

Before getting into the details of the formal consequences of $(\mathcal{I})$ in the context of Strevens's account, we begin with a cautionary remark. Bayesian readers should be rather suspicious when something like $(\mathcal{I})$ is applied in this kind of context. Basically, the legitimacy of using $(\mathcal{I})$ depends on whether logically weakening (or strengthening) the evidence in Q–D cases can have drastic confirmation-theoretic effects. There are many examples in the literature that illustrate the significant effects that can be caused by the logical weakening of evidence. In general, Bayesian confirmation is quite sensitive to such weakenings. That this is the case (in a general, abstract sense) has been known for some time (see Carnap [1950, §57]). And, that there are important real examples of this phenomenon in statistical inference has also been known for quite a while. For instance, I.J. Good [1960] discusses the effect of evidential weakenings in connection with Hempel's ravens paradox (under the rubric of 'stoogian' evidence reports which ignore certain features of the data). As we will see below, it is this very phenomenon that underlies the problems with Strevens's use of $(\mathcal{I})$ in this context.

3 Indications that $(\mathcal{I})$ Oversimplifies Q–D

For the most part, Strevens speaks about the *posterior probabilities* of H and A (on evidence E) in his discussion of the relative degrees of confirmation E provides for H *vs* A. However, Strevens sometimes (*e.g.*, p. 526) uses the Bayesian *ratio measure* of the degree to which y *incrementally* confirms x: $r(x, y) = \frac{\Pr(x \mid y)}{\Pr(x)}$. This measure of confirmation (or something ordinally equivalent to it) has been defended by various Bayesians (*e.g.*, Peter Milne [1996]). While we think this measure and the posterior probability measure of degree of confirmation are both inadequate (see Eells and Fitelson [2002] and Fitelson [2002] for several arguments against r), we will not (for now) criticize Strevens on these grounds. What's most important here is how Strevens uses his idealization $(\mathcal{I})$ to simplify the expressions $\mathfrak{c}(H, E)$ and $\mathfrak{c}(A, E)$ (for *whichever*

Bayesian measure of confirmation $\mathfrak{c}$ you prefer). As it turns out, ($\mathcal{I}$) is much more powerful in this respect than Strevens's discussion seems to suggest. Strevens argues that ($\mathcal{I}$) does not oversimplify the comparison of $\mathfrak{c}(H, E)$ and $\mathfrak{c}(A, E)$. In particular, Strevens suggests that — even if we assume ($\mathcal{I}$) — the comparative relationship between $\mathfrak{c}(H, E)$ and $\mathfrak{c}(A, E)$ will depend not only on the prior probabilities of H and A, but also on various conditional probabilities such as $\Pr(A \mid H)$ and $\Pr(H \mid A)$ (see Strevens's plots and discussion on pages 526–527). The proof of the following theorem shows that this is not the case, for *any* of the Bayesian measures of confirmation that we have seen in recent use, *including* the posterior probability measure, and all contemporary incremental measures, including r.[2]

Theorem 1. *Assuming* ($\mathcal{I}$), $\mathfrak{c}(H, E) \geq \mathfrak{c}(A, E)$ *if and only if* $\Pr(H) \geq \Pr(A)$ *(whether* $\mathfrak{c}$ *is the posterior probability measure, or any incremental measure such as* r*).*

That is, enforcing ($\mathcal{I}$) reduces the comparison of $\mathfrak{c}(H, E)$ and $\mathfrak{c}(A, E)$ to a simple comparison of the priors $\Pr(H)$ and $\Pr(A)$, which renders the evidence E *irrelevant* (*a posteriori*) to this comparison.[3] It seems clear to us that, in the original Q–D problem, the mere relationship between the priors of H and A should not *by itself* determine the relative support that E provides for H *vs* A. We take Quine and Duhem to be asking the following question: In cases where $H \,\&\, A$ *entails* $\neg E$, can the evidence E differentially confirm H *vs* A — *a posteriori* — and, if so, *how*? This is *not* a question about the relative *a priori* plausibilities of H *vs* A, but rather a question about the *a posteriori* confirmational power of E to discriminate between H and A when $H \,\&\, A$ entails $\neg E$. What Theorem 1 shows is that Strevens's assumption ($\mathcal{I}$) makes it *impossible* to use Bayesianism to provide an affirmative answer to this question (or to provide an explanation of how an affirmative answer is possible). *Pace* Strevens, we think an affirmative answer to the Q–D question *is* possible within a Bayesian framework, and we take this ability of Bayesianism (to do what deductive accounts of confirmation cannot seem to do) to be a virtue. Here is a simple example that makes our worry about ($\mathcal{I}$) vivid.

Consider a lottery involving 11 tickets. Let H be the claim that either the winning ticket will be among tickets #1–#5 or it will be ticket #11. Let A be the claim that the winning ticket will be among tickets #4–#9, and let E be the claim that either the winning ticket is among tickets #1–#3 or it is among tickets #6–#9. It seems to us intuitively clear that E favors A over H in this example, since $\mathfrak{c}(A, E) > \mathfrak{c}(H, E)$, for *any* Bayesian measure of confirmation $\mathfrak{c}$, including the posterior probability measure and all incremental measures. Moreover, this is a Q–D case, in which $H \,\&\, A$ entails $\neg E$, $\Pr(A) = \Pr(H)$, E confirms A[4] (in the incremental sense), and E disconfirms H.[5]

[2] See the Appendix for proofs of all theorems.

[3] Other, more sophisticated probabilistic analyses of Q–D do not have this consequence. See Dorling [1979], Howson and Urbach [1993], and Sober [2004] for examples of alternative approaches to Q–D that are immune to our criticisms of Strevens.

[4] If we assume ($\mathcal{I}$), then *E cannot confirm either H or A (individually) in an incremental sense*, since ($\mathcal{I}$) also entails that $\Pr(H \mid E) \leq \Pr(H)$ and $\Pr(A \mid E) \leq \Pr(A)$. Strevens [2001, p. 526] seems to be aware of, but unmoved by, this consequence of ($\mathcal{I}$). Our example indicates that this is another oversimplification of Q–D.

[5] We omit the simple calculations which verify these claims about our simple Q–D example. Further simple calculations reveal that our example has some other useful features as well. For instance, in our example, H and A are *not* probabilistically independent. Therefore, Dorling's [1979] assumption that H and A are probabilistically independent is *not necessary* for an affirmative Bayesian answer to the Q–D question.

So, H and A have *equal a priori* probability, *but* E confirms A more strongly than E confirms H. We take it that this example suffices to show that Bayesianism is capable of answering the Q–D question in the affirmative, and explaining how an affirmative answer is possible. Unfortunately, because it assumes ($\mathcal{I}$), Strevens's account cannot do so. The problem here is that if we look *only* at the part of E that is *equivalent to* $\neg(H \,\&\, A)$, we cannot explain why E favors A over H in this example (or others like it), since to do so would require an appeal to A's having greater *a priori* plausibility than H. But, *A doesn't have* greater *a priori* plausibility than H here, so no explanation based on Strevens's ($\mathcal{I}$) is forthcoming.

Our example suggests that ($\mathcal{I}$) oversimplifies the original Q–D problem. But, the more general point is that to look *only* at the part of E that *falsifies H & A* (as Strevens recommends) is to ignore precisely that part of E that would allow a Bayesian to answer the challenge of Quine and Duhem. In general, probabilistic approaches to confirmation work best when they take account of the totality of the evidence, and not some censored version of it. What Strevens's paper makes clear is that taking into account the full content of E [*i.e.*, not ignoring the part of E that goes beyond $\neg(H \,\&\, A)$] is essential to resolving the Q–D problem within a probabilistic framework.

4 Strevens's Argument for the Legitimacy of ($\mathcal{I}$)

Our discussion so far suggests that ($\mathcal{I}$) is too strong a 'simplifying idealization' to yield a useful general Bayesian rendition and analysis of the original Q–D problem. But, these considerations do not furnish a diagnosis of the source of the problem in the development and justification of the account. Strevens presents an argument, which is supposed to show that ($\mathcal{I}$) is not an undue oversimplification of the Q–D problem. First, Strevens returns to the original Q–D problem, in which $E \vDash \neg(H \& A)$, but $\neg(H \& A) \nvDash E$ (*i.e.*, in which ($\mathcal{I}$) is relaxed). Then, he aims (*i*) to identify a condition C under which the "partial posterior" $\Pr(H \mid \neg(H \,\&\, A))$ is *approximately equal to* the "full posterior" $\Pr(H \mid E)$, and then (*ii*) to argue that this condition C obtains in the "interesting" Q–D cases. Here's a concise exposition of Strevens's strategy for justifying idealization ($\mathcal{I}$).

1. Assume (as in the original Q–D problem) *only* that $E \vDash \neg(H \,\&\, A)$.

2. $\therefore \Pr(H \mid \neg(H \,\&\, A)) = \Pr(H \mid E)$ iff $\Pr(E \mid H \,\&\, \neg A) = \Pr(E \mid \neg(H \,\&\, A))$.

3. $\therefore \Pr(H \mid \neg(H \,\&\, A)) \approx \Pr(H \mid E)$ if $\Pr(E \mid H \,\&\, \neg A) \approx \Pr(E \mid \neg(H \,\&\, A))$.

4. $\Pr(E \mid H \,\&\, \neg A) \approx \Pr(E \mid \neg(H \,\&\, A))$ in all "interesting" Q–D cases.

5. $\therefore \Pr(H \mid \neg(H \,\&\, A)) \approx \Pr(H \mid E)$ in all "interesting" Q–D cases.

---

6. $\therefore$ It is legitimate to make use of ($\mathcal{I}$) in a Bayesian account of Q–D.

The basic idea here is to establish that $\Pr(H \mid \neg(H \,\&\, A)) \approx \Pr(H \mid E)$ holds in all "interesting" Q–D cases, and as a result, that there is no real harm in the idealization that E is *no stronger than* $\neg(H \& A)$, since this won't radically (or unfairly) affect our judgments concerning the posterior probability of H (and, presumably, by symmetry, the same should hold for A), at least, not in the "interesting" or "important" Q–D cases (*e.g.*, cases in which Strevens's so-called "Newstein Effects" are absent).

We think there are various ways this argument goes wrong. Specifically, we think that premises (**3**), (**4**), and (**5**) of the above argument are all false, and that the $\therefore$'s which precede (**3**), (**5**), and (**6**) are fallacious. We will not discuss the problems with premise (**4**) presently (these problems, which involve Strevens's discussion of what he calls "Newstein Effects", are rather subtle and would require more space to flesh out). Our focus will be on steps (**3**) and (**6**), which are fallacious. As for the rest of the argument, premises (**1**) and (**2**) are beyond reproach. (**2**) is a theorem (given assumption (**1**)), and (**1**) is a basic assumption of Q–D. The following theorem exposes what we take to be one of two key missteps in the above argument [in the step from (**2**) to (**3**)]:

Theorem 2. *Assuming* $E \vDash \neg(H \,\&\, A)$*, it is possible for* $\Pr(E \mid H \,\&\, \neg A)$ *and* $\Pr(E \mid \neg(H \,\&\, A))$ *to be* arbitrarily close*, while* at the same time $\Pr(H \mid \neg(H \,\&\, A))$ *and* $\Pr(H \mid E)$ *are arbitrarily far apart [despite the fact that* "$\Pr(H \mid \neg(H \,\&\, A)) = \Pr(H \mid E)$ *iff* $\Pr(E \mid H \,\&\, \neg A) = \Pr(E \mid \neg(H \,\&\, A))$" *does follow from* $E \vDash \neg(H \,\&\, A)$*, as Strevens shows].*

What this shows is that despite the fact that (**2**) is a theorem (assuming (**1**)), (**3**) does not follow. In fact, even in Q–D cases, the likelihoods $\Pr(E \mid H \,\&\, \neg A)$ and $\Pr(E \mid \neg(H \,\&\, A))$ can be *as close as one likes*, while *at the same time* the "partial" and "full" posteriors $\Pr(H \mid \neg(H \,\&\, A))$ and $\Pr(H \mid E)$ are *as far apart as one likes*. So, the claim that the "partial" and "full" posteriors can be assumed to be "close" in the "interesting" Q–D cases, because the likelihoods in question are "close" in such cases is false (no matter how liberally we interpret "$\approx$").

The second misstep in the above argument is the step from (**5**) to (**6**). Premise (**5**) is relevant to the confirmational comparison $\mathfrak{c}(H, E) \geq \mathfrak{c}(A, E)$ if one assumes that the posterior probability is an adequate Bayesian measure of degree of confirmation. But, contemporary Bayesians think of confirmation in terms of degree of probabilistic relevance, not in terms of degree of conditional probability. As such, contemporary Bayesians use relevance measures of confirmation that disagree radically with the posterior probability when it comes to comparative claims like $\mathfrak{c}(H, E) \geq \mathfrak{c}(A, E)$. Various relevance measures have been proposed and defended in the literature on confirmation theory (see Fitelson [1999] for a brief survey). Some of these measures will sanction the step from (**5**) to (**6**) in Q–D cases, and some will not. The measure we favor is the likelihood ratio measure $l(H, E) = \frac{\Pr(E \mid H)}{\Pr(E \mid \neg H)}$. See Fitelson [2001] for various reasons to favor this measure over other relevance measures. As it turns out, even if (**5**) is true (in a Q–D case), it does not follow that $l(H, \neg(H \,\&\, A)) \approx l(H, E)$.[6] We think this makes the step from (**5**) to (**6**) fallacious (at least, in the absence of reasons not to use relevance measures of confirmation, like l or r) from the point of view of contemporary Bayesian confirmation theory. To wit:

Theorem 3. *Assuming* $E \vDash \neg(H \,\&\, A)$*, it is possible for* $\Pr(H \mid \neg(H \,\&\, A))$ *and* $\Pr(H \mid E)$ *to be* arbitrarily close*, while* at the same time $l(H, \neg(H \,\&\, A))$ *and* $l(H, E)$ *are arbitrarily far apart (where* $l(H, E) = \frac{\Pr(E \mid H)}{\Pr(E \mid \neg H)}$ *is the likelihood ratio measure of the degree to which E incrementally confirms H).*

We think that further analysis will reveal other problems with Strevens's strategy for justifying his use of ($\mathcal{I}$) (*e.g.*, concerning his premise (**4**)). But, those worries would require a separate paper to articulate properly. So, we now conclude our discussion.

[6]This step is fallacious for many other relevance measures of confirmation, including r (proof omitted).

Acknowledgments

Thanks to Alan Hájek, Jim Hawthorne, Jim Joyce, Elliott Sober, Michael Strevens, and an anonymous referee of this journal for useful comments and suggestions.

Branden Fitelson
Department of Philosophy
University of California
Berkeley, CA 94720–2390
USA
branden@fitelson.org

Andrew Waterman
andrew.waterman@stanford.edu

Appendix: Proof of Theorems

Theorem 1. *Assuming* ($\mathcal{I}$), $\mathfrak{c}(H, E) \geq \mathfrak{c}(A, E) \Leftrightarrow \Pr(H) \geq \Pr(A)$ *(whether* $\mathfrak{c}$ *is the posterior probability measure, or an incremental measure such as r).*

Proof. We will prove this theorem with a combination of algebraic and axiomatic technique. All we need is a 2-event probability space, containing the events H and A. Then, we have four elementary probabilities in the space, each of which is on $[0, 1]$, and the sum of which is 1. These four elementary probabilities are as follows:

$\Pr(H \,\&\, \neg A) = a$	$\Pr(H \,\&\, A) = b$
$\Pr(\neg H \,\&\, A) = c$	$\Pr(\neg H \,\&\, \neg A) = d$

Now, the proof for the ratio measure r proceeds as follows.[7]

$$
\begin{aligned}
\Pr(H) \geq \Pr(A) &\Leftrightarrow \Pr(H \,\&\, \neg A) + \Pr(H \,\&\, A) \geq \Pr(H \,\&\, A) + \Pr(\neg H \,\&\, A)\\
&\Leftrightarrow a + b \geq b + c\\
&\Leftrightarrow a \geq c\\
&\Leftrightarrow ab + ac \geq ac + bc\\
&\Leftrightarrow a(b + c) \geq (a + b)c\\
&\Leftrightarrow \frac{a}{(a + b)(a + c + d)} \geq \frac{c}{(a + c + d)(b + c)}\\
&\Leftrightarrow \frac{\Pr(H \mid \neg(H \,\&\, A))}{\Pr(H)} \geq \frac{\Pr(A \mid \neg(H \,\&\, A))}{\Pr(A)}\\
&\Leftrightarrow r(H, \neg(H \,\&\, A)) \geq r(A, \neg(H \,\&\, A))
\end{aligned}
$$

[7] We assume without comment throughout the Appendix that the salient conditional probabilities are defined. This amounts to assuming that the salient elementary events in the probability models discussed have non-zero probabilities, where needed, to ensure that we are not dividing by zero anywhere. Also, along the same lines (and strictly speaking), the $\Rightarrow$ direction of Theorem 1 requires the additional assumption that $\Pr(H \,\&\, A) = b \neq 0$. But, in the context of Q–D, this is *always* assumed, since otherwise the theory $T = H \,\&\, A$ would be unconfirmable (and undisconfirmable) by *anything*, including its logical consequences.

Therefore, by ($\mathcal{I}$), $\Pr(H) \geq \Pr(A) \Leftrightarrow r(H, E) \geq r(A, E)$. We omit the (similar) proofs for the other Bayesian measures of confirmation currently being used and defended, *e.g.*, the posterior probability measure $p(x, y) = \Pr(x \mid y)$, the difference measure $d(x, y) = \Pr(x \mid y) - \Pr(x)$, the likelihood ratio measure $l(x, y) = \frac{\Pr(y \mid x)}{\Pr(y \mid \neg x)}$, and the normalized difference measure $s(x, y) = \Pr(x \mid y) - \Pr(x \mid \neg y)$. See Fitelson [1999] for a contemporary survey of (incremental) Bayesian measures of confirmation. ❑

Theorem 2. *Assuming* $E \vDash \neg(H \& A)$*, it is possible for* $\Pr(E \mid H \& \neg A)$ *and* $\Pr(E \mid \neg(H \& A))$ *to be* arbitrarily close*, while* at the same time $\Pr(H \mid \neg(H \& A))$ *and* $\Pr(H \mid E)$ *are arbitrarily far apart [despite the fact that "*$\Pr(H \mid \neg(H \& A)) = \Pr(H \mid E)$ *iff* $\Pr(E \mid H \& \neg A) = \Pr(E \mid \neg(H \& A))$*" does follow from* $E \vDash \neg(H \& A)$*, as Strevens shows].*

Proof. We prove this theorem by describing an algorithm for generating probability models in which (*i*) $E \vDash \neg(H \& A)$, (*ii*) $|\Pr(E \mid H \& \neg A) - \Pr(E \mid \neg(H \& A))| \leq \epsilon$ for arbitrary $\epsilon \in (0, \frac{1}{2})$, and (*iii*) $|\Pr(H \mid \neg(H \& A)) - \Pr(H \mid E)| = 1 - \epsilon$. There are many ways to do this. Any one of these would be sufficient to establish the theorem. Here is the first such algorithm we found.[8] We need an algorithm for generating a 3-event probability model, containing the events H, A, and E, such that (*i*)–(*iii*) above are all satisfied. Such models will have eight elementary probabilities, each of which is on $[0, 1]$, and the sum of which is 1. Our algorithm is as follows.[9]

Step 1. Pick an $\epsilon \in (0, \frac{1}{2})$, which is as small as you like.
Step 2. Assign the following numbers to the eight elementary probabilities in the space:

$\Pr(H \& \neg E \& \neg A) = -2\,\epsilon^3 + \epsilon^4 + \frac{\epsilon}{1+\epsilon}$	$\Pr(H \& E \& \neg A) = \frac{\epsilon^3+\epsilon^4-\epsilon^5}{1+\epsilon}$
$\Pr(\neg H \& E \& \neg A) = -\left(\frac{\epsilon^5\left(-1+\epsilon+\epsilon^2\right)}{1+\epsilon}\right)$	$\Pr(H \& \neg E \& A) = \epsilon^2$
$\Pr(H \& E \& A) = 0$	$\Pr(\neg H \& E \& A) = \epsilon^6$
$\Pr(\neg H \& \neg E \& A) = -\left(\epsilon\left(-1 + \epsilon + \epsilon^5\right)\right)$	$\Pr(\neg H \& \neg E \& \neg A) = -\epsilon + \epsilon^3 - \epsilon^4 + \epsilon^6 + \frac{1}{1+\epsilon}$

This procedure will always generate a probability model in which (*i*)–(*iii*) obtain.[10] ❑

Theorem 3. *Assuming* $E \vDash \neg(H \& A)$*, it is possible for* $\Pr(H \mid \neg(H \& A))$ *and* $\Pr(H \mid E)$ *to be* arbitrarily close*, while* at the same time $l(H, \neg(H \& A))$ *and* $l(H, E)$ *are arbitrarily*

[8] We do not claim that the models generated by this algorithm will correspond to "interesting" Q–D cases. However, our goal here is simply to establish the *formal* fallaciousness of Strevens's reasoning. We view this algorithm as a way to shift the burden of proof back onto Strevens. It is now up to him to explain why the "interesting" Q–D cases are *not* susceptible to the formal fallacy illustrated by our algorithm.

[9] We assume, for simplicity, that the probability models in question are *regular*. We could weaken this assumption, and specify that *in addition to* $\Pr(H \& E \& A) = 0$, one must *also* make sure that $E \vDash \neg(H \& A)$. This is a minor technical detail that arises from the distinction between probability zero and logical falsity.

[10] This claim is easily verified by algebraic computation (details omitted). More generally, to get Theorem 2 all we need is to have $\Pr(H \& \neg E \& \neg A) \approx 1$, $\Pr(E) \approx 0$, and $\Pr(H \& E \& A) = 0$. On request, the authors will supply a *Mathematica* 5.0 notebook which verifies the (mathematical) claims made in this Appendix.

far apart (where $l(H, E) = \frac{\Pr(E \mid H)}{\Pr(E \mid \neg H)}$ *is the likelihood ratio measure of the degree to which E incrementally confirms H).*

Proof. We prove this theorem by describing an algorithm for generating probability models in which (*iv*) $E \vDash \neg(H \,\&\, A)$, (*v*) $|\Pr(H \mid \neg(H \,\&\, A)) - \Pr(H \mid E)| = \epsilon$ for arbitrary $\epsilon \in (0, \frac{1}{4})$, and (*vi*) $l(H, E) = \epsilon \cdot l(H, \neg(H \,\&\, A))$. The algorithm is as follows:

Step 1. Pick an $\epsilon \in (0, \frac{1}{4})$, which is as small as you like.
Step 2. Assign the following numbers to the eight elementary probabilities in the space:

$$\Pr(H \,\&\, \neg E \,\&\, \neg A) = \frac{-1+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}-\epsilon\left(-1+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}+\epsilon\left((-1+\epsilon)\,\epsilon+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}\right)\right)}{2\,(-1+\epsilon)}$$

$$\Pr(H \,\&\, E \,\&\, \neg A) = \frac{-\left(\epsilon\left(1+(-2+\epsilon)\,\epsilon-\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}\right)\right)}{2\,(-1+\epsilon)}$$

$$\Pr(\neg H \,\&\, E \,\&\, \neg A) = \frac{-\left(\epsilon\left(1+(-2+\epsilon)\,\epsilon+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}\right)\right)}{2\,(-1+\epsilon)}$$

$$\Pr(H \,\&\, \neg E \,\&\, A) = \epsilon^2$$

$$\Pr(H \,\&\, E \,\&\, A) = 0$$

$$\Pr(\neg H \,\&\, E \,\&\, A) = \epsilon^2$$

$$\Pr(\neg H \,\&\, \neg E \,\&\, A) = \epsilon - 2\,\epsilon^2$$

$$\Pr(\neg H \& \neg E \& \neg A) = \frac{-\left(1+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}-\epsilon\left(5+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}+\epsilon\left(-6+\epsilon+\epsilon^2+\sqrt{1+\epsilon\left(-4+2\,\epsilon+\epsilon^3\right)}\right)\right)\right)}{2\,(-1+\epsilon)}$$

This procedure will always generate a probability model in which (*iv*)–(*vi*) obtain.[11] That is sufficient to establish the theorem, and to complete the Appendix. ❑

References

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Dorling, J. (1979). Bayesian personalism, the methodology of scientific research programmes, and Duhem's problem. *Studies in History and Philosophy of Science 10*(3), 177–187.

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

[11] It is interesting to note that Dorling [1979, p. 178] makes a similar mistake, when he claims that, as long as we have a "sufficiently good approximation of" $\Pr(H)$, we will automatically have a "sufficiently good approximation of" $\mathfrak{c}(H, E)$, provided that we know $\Pr(E \mid H)$ and $\Pr(E)$ with perfect precision. This is false if $\mathfrak{c}$ is taken to be the likelihood-ratio measure (it's false for some other choices of measure as well). This sort of "approximation fallacy" has appeared in various forms in the literature (we know of several other recent examples). But, to our knowledge, it has received almost no attention. In general, one must be very careful about using such "approximate" reasoning in non-linear spaces such as Kolmogorov probability models.

Eells, E. and B. Fitelson (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies 107*(2), 129–142.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science 66*, S362–S378.

Fitelson, B. (2001). *Studies in Bayesian Confirmation Theory*. Ph. D. thesis, University of Wisconsin–Madison (Philosophy).

Fitelson, B. (2002). Putting the irrelevance back into the problem of irrelevant conjunction. *Philosophy of Science 69*(4), 611–622.

Good, I. (1960). The paradox of confirmation. *The British Journal for the Philosophy of Science 11*(42), 145–149.

Howson, C. and P. Urbach (1993). *Scientific Reasoning: The Bayesian Approach*. La Salle: Open Court.

Milne, P. (1996). $\log[p(h/eb)/p(h/b)]$ is the one true measure of confirmation. *Philosophy of Science 63*, 21–26.

Sober, E. (2004). Likelihood, model selection, and the Duhem-Quine problem. To appear in the *Journal of Philosophy*.

Strevens, M. (2001). The Bayesian treatment of auxiliary hypotheses. *British Journal for the Philosophy of Science 52*(3), 515–537.