Commentary on ‘The Probability of the Evidence’ by Sherrilyn Roush

James Justus
Program in the History and Philosophy of Science
University of Texas at Austin
justus.phil@mail.utexas.edu

Roush argues that $p(e)$ should be close but not equal to 1. She also criticizes the claim that a low value of $p(e)$ best represents the intuition that surprising evidence is more confirmatory than nonsurprising evidence. She argues:

(i) high values of $p(e)$ and the likelihood ratio (LR) put a lower bound on $p(h|e)$;
(ii) the value of $p(e)$ is, intuitively, close to 1 and this value is required by Bayesian conditionalization on e;
(iii) “a scheme in which we determine high values for $p(e)$ and for the likelihood ratio in order to determine that e is evidence for h makes sense of the familiar practice of eliminative reasoning in science and elsewhere” (p. 1).

1. Point (i).
Roush shows that Bayes’ theorem is equivalent to:

$$\text{(BT}'\text{)} \quad p(h\,|\,e) = \frac{LR - \dfrac{p(e\,|\,h)}{p(e)}}{LR - 1};$$

in which $\mathrm{LR} = \dfrac{p(e\,|\,h)}{p(e\,|\,\neg h)}$ and $p(h)$ does not explicitly occur. A graph of $p(h|e)$ as a function of $p(e|h)$, $p(e)$, and LR shows that, in general, for sufficiently high values of $p(e)$ and LR, lower bounds on these terms set a lower bound on $p(h|e)$. Roush’s graphs, however, may be misleading since the axes ($p(e|h)$, $p(e)$, and LR) are not independent. The conclusion, fortunately, can be established formally:

$$p(h\,|\,e) = \left(\frac{LR - \dfrac{p(e\,|\,h)}{p(e)}}{LR - 1} \right) > \left(\frac{LR - \dfrac{p(e\,|\,h)}{p(e)}}{LR} \right) = \left(1 - \frac{p(e\,|\,h)}{p(e)} \frac{1}{LR} \right) \geq \left(1 - \frac{1}{p(e)} \frac{1}{LR} \right).$$

Roush then suggests:

> A lower bound on the posterior probability of the hypothesis can only be good, because knowing this means knowing something about whether the hypothesis is true, which we do not get from positive relevance or even from high LR. If e is positively relevant to h, that is, $p(h|e)>p(h)$, this does not put a lower bound on the posterior probability of h unless we have a lower bound for $p(h)$…e may be positively relevant to h while giving us no good reason to believe h (p. 4).

Having a good reason to believe h for Roush means $p(h)>0.5$. Alone, the quote does not show, nor does Roush intend it to show, a virtue of her approach over positive relevance. Roush does suggest that one disadvantage of the positive relevance view is that it is hard to know when $p(h)>0.5$ and $p(h|e)>p(h)$, and in a previous draft of this paper Roush claimed one virtue of (BT′) is that it does not require evaluation of $p(h)$ or $p(\neg h)$, which she no longer defends.

Roush's claim, however, may have a better defense. Roush could argue that, even though $p(e|h)$ and $p(e|\neg h)$ (the numerator and denominator of LR) are partially defined by $p(h)$ and $p(\neg h)$ on standard axiomatizations of probability theory, determining the values of $p(e|h)$ and $p(e|\neg h)$, as a matter of scientific practice, is less problematic than evaluating $p(h)$ and $p(\neg h)$. In fact, one of the primary purposes of scientific methodology is to pinpoint the value of $p(e|h)$ –the degree a hypothesis h predicts the data e– as precisely as possible. It is unclear, however, that scientists are similarly concerned with determining $p(e|\neg h)$. In practice the concern is usually with a small set $\{p(e|h_i)$: $h_i \in \mathbf{H}$, $i=1,...,n\}$, where **H** is a set of hypotheses considered plausible but probably not exhaustive of all possible hypotheses that account for e. Methodological questions usually focus on distinguishing $p(e|h_r)$ and $p(e|h_p)$ for some r and p, $r \neq p$. They usually do not, nor does it seem they usually need to, focus on determining the corresponding likelihoods $p(e|\neg h_r)$ and $p(e|\neg h_p)$ or on determining the "catchall" likelihood: $p(e|\neg[h_1 \vee h_2 \vee \ldots h_n])$. Of course, scientific methodology involves a diverse set of contexts and complicated methods of testing so there may be cases in which evaluation of $p(e|h_i)$ and $p(e|\neg h_i)$ for each h_i of some set of hypotheses **H** is appropriate or even necessary. This alone, however, does not seem to demonstrate a clear advantage of Roush's approach over positive relevance.

2. **Point (ii)**: **The Intuitiveness of High** $\boldsymbol{p(e)}$.

Besides LR and $p(e|h)$, according to (BT′) $p(h|e)$ is also a function of $p(e)$, which Roush argues should be close but not equal to 1. On a personalist interpretation of probability, $p(e)$ is an agent's degree of belief in e. Specifically, it does not, according to Roush, necessarily represent the "expectedness" of e. Yet a personalist interpretation, Roush believes, requires that e is evidence iff $p(e)$ is close to 1. This is incompatible with the claim that a low value of $p(e)$ best represents the intuition that surprising evidence is more confirmatory than nonsurprising evidence.

In support, Roush suggests, "no one takes a deeply surprising occurrence as evidence of anything until he satisfies himself that it did indeed occur, but then he has a high degree of belief that it occurred, and $p(e)$ must be high" (p. 8). Roush's account of Rutherford's ultimate belief that back-scattering occurs at a particular rate illustrates her point:

> Back-scattering of alpha particles from thin gold foil was deeply surprising to Ernest Rutherford, but he wouldn't have taken the statement, e_s, that back-scattering occurred at a certain rate, as evidence of the existence of a nucleus in the atom unless he had done some checking to reassure himself that e_s was true (p. 8).

Scientists must establish that reports of evidence are veridical, that what they take to be evidence they believe to be true. Scientists may still believe, however, that some particular evidence *e* is surprising even after it has been rigorously demonstrated to occur. It is plausible, if not probable, that Rutherford would agree that back-scattering remains surprising even after he had conclusively verified the existence of the phenomenon with repeated experiments. Given that the surprisingness of *e* may remain after one is certain that *e* occurs, the claim that $p(e)$ should represent the "expectedness" of *e* in Bayes' Theorem, which would have a low value in such a case, cannot be dismissed as mistaken. Specifically, it remains plausible that the Bayesian should calculate $p(h|e)$ based on the value of $p(e)$ understood as the expectedness of *e*, not the degree of belief in *e* once it has been rigorously shown to occur. What Roush has shown, however, it that there is a cost for this commitment. It abandons the straightforward interpretation of $p(e)$ as the degree of belief in *e*.

3. Point (ii): Bayesian Conditionalization and High *p*(*e*).

Roush quotes Howson and Urbach's (1993, p. 99) characterization of strict Bayesian conditionalization:

> When your degree of belief in *e* goes to 1, but no stronger proposition also acquires probability 1, set $p'(a) = p(a|e)$ for all *a* in the domain of *p*, where *p* is your probability function immediately prior to the change.

and claims it says, "your degree of belief in *e* approaching 1 is a sufficient condition for conditionalizing on *e*" (p. 9) as long as no stronger proposition also acquires probability 1. Since, Roush also argues (p. 9), Bayesians believe that strict conditionalization is the only defensible rule of updating, a high value of $p(e)$ is also a necessary condition for *e* to be evidence. The assumption underlying both of these claims is that *e* is evidence iff it can be the basis for strict conditionalization. The underlying (and compelling) intuition is that only a high value of $p(e)$ *justifies* conditionalization.

3.1. Extrapolation from Point (i) and (ii).

If Roush's argument that *e* is evidence iff $p(e)$ is high is sound, she has shown that *e* can be the basis for strict Bayesian conditionalization iff $p(e)$ is high. But if $p(e)$ is high it will little confirm *h* on the positive relevance view of confirmation. Roush's argument, therefore, poses a dilemma to personalist Bayesians. If, as personalists, they interpret $p(e)$ as degree of belief in *e*, they cannot adopt positive relevance as the best measure of confirmation since *e* will little confirm *h* on this measure for any *e* and *h*. If, on the other hand, they believe that a low value of $p(e)$ best represents the intuition that surprising evidence is more confirmatory and adopt the positive relevance view of confirmation, they must abandon the straightforward interpretation of $p(e)$ as the degree of belief in *e*. Roush's argument thereby reinforces a point made in the question and answer period that a rule for belief updating (*e.g*. strict Bayesian or Jeffrey conditionalization) probably will not provide a good measure of confirmation.

4. Point (iii).

Roush claims that finding values of $p(e)$ close but not equal to 1 and high values of LR "looks like" (p. 10) or "makes sense of" (p. 1) the eliminative reasoning used by scientists. Given a set of probably not exhaustive hypotheses $\{h_i: i=1,...,n\}$, eliminative reasoning involves finding, "what could be responsible for *e*'s occurrence" (p. 10).

Roush claims that a hypothesis h_e can be eliminated if $p(h_e)$ or $p(e|h_e)$ is very low. Roush then states:

> findings of this sort are what we need in order to evaluate the likelihood of *e* on all possible hypotheses ... [this is what] we take into account when we try to determine what is responsible for the occurrence of *e*, and accordingly, what *e* is evidence for. Thus, evaluation of LR corresponds to the eliminative reasoning scientists engage in when they try to determine what *e* is evidence for.

It is unclear what these remarks demonstrate. First, even if evaluation of LR corresponds to scientific eliminative reasoning, this seems independent of the issue of the value of $p(e)$. It does not, specifically, establish that $p(e)$ is close but not equal to 1, or that LR is sufficiently high to (usually) set a lower bound on $p(h|e)$ that is >0.5. Second, that determining LR may correspond to scientific eliminative reasoning does not alone justify the claim that LR (and thereby BT′) has a special normative or epistemic status. Actual scientific methodology involves numerous kinds of reasoning, some of which may correspond to other confirmation measures, and some of which may be more justified than others. Without further argumentation, the potential correspondence Roush claims does not seem to justify any significant normative conclusions.

Reference

Howeson, C. and Urbach, P. (1993), *Scientific Reasoning: The Bayesian Approach*. 2nd edition. Chicago: Open Court Publishing.