

## Does Roush show that evidence should be probable?

Damien Fennell · Nancy Cartwright

Received: 9 July 2008 / Accepted: 16 March 2009 / Published online: 21 April 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** This paper critically analyzes Sherrilyn Roush’s (Tracking truth: knowledge, evidence and science, 2005) definition of evidence and especially her powerful defence that in the ideal, a claim should be probable to be evidence for anything. We suggest that Roush treats not one sense of ‘evidence’ but three: relevance, leveraging and grounds for knowledge; and that different parts of her argument fare differently with respect to different senses. For relevance, we argue that probable evidence is sufficient but not necessary for Roush’s own two criteria of evidence to be met. With respect to grounds for knowledge, we agree that high probability evidence is indeed ideal for the central reason Roush gives: When believing a hypothesis on the basis of  $e$  it is desirable that  $e$  be probable. But we maintain that her further argument that Bayesians need probable evidence to warrant the method they recommend for belief revision rests on a mistaken interpretation of Bayesian conditionalization. Moreover, we argue that attempts to reconcile Roush’s arguments with Bayesianism fail. For leveraging, which we agree is a matter of great importance, the requirement that evidence be probable suffices for leveraging to the probability of the hypothesis if either one of Roush’s two criteria for evidence are met. Insisting on both then seems excessive. To finish, we show how evidence, as Roush defines it, can fail to track the

---

D. Fennell (✉)  
Centre for Philosophy of Natural and Social Science, London School of Economics,  
London WC2A 2AE, UK  
e-mail: d.j.fennell@lse.ac.uk

N. Cartwright  
Department of Philosophy, Logic and Scientific Method, London School of Economics,  
London WC2A 2AE, UK

N. Cartwright  
Department of Philosophy,  
University of California San Diego, San Diego 92093-0119, USA

hypothesis. This can be remedied by adding a requirement that evidence be probable, suggesting another rationale for taking probable evidence as ideal—but only for a grounds-for-knowledge sense of evidence.

**Keywords** Evidence · Philosophy of science · Probability · Surprising evidence

## 1 Introduction

Evidence has always been a central topic in the philosophy of science. But with debates raging throughout the US and the UK about what counts as evidence for evidence-based policy, the topic has become one of immediate practical importance. This makes very timely Sherrilyn Roush's recent efforts to characterize evidence.

Roush presents her account near the end of her detailed development of a tracking theory of knowledge, now one of the central topics in epistemology. Even if tracking is not the correct or the only good account of knowledge, it would certainly be a plus to have a theory of evidence for which it could be assumed that anything that counts as evidence is a source of tracking knowledge. We shall here look into the question of how her theory of evidence dovetails with her tracking view of knowledge.

Our primary topic, however, is another important claim that Roush defends. Roush claims for her account the special virtue that it “explains why better evidence makes knowledge more probable” (Roush 2005, p. 185). This seems a truism: What we mean by better evidence for  $h$  is evidence that makes  $h$  more probable. But this is not what Roush means. For her, ‘better evidence’ is evidence that has a higher probability. If  $e$  is to be evidence for anything, she maintains, it is ideal that  $P(e)$  be high.

This conclusion matters, and not just for the philosophical explication of the concept of evidence. Now that evidence-based policy is widely mandated, guides outlining what counts as evidence for policy effectiveness proliferate.<sup>1</sup> Roush's demand for high  $P(e)$  is at their heart. The guides offer schemes that rank methods for producing evidence according to the degree of certainty that the method confers on the conclusions it produces: High-quality evidence claims are claims produced by methods that make it likely that those claims are true, claims  $e$  for which  $P(e)$  is high,<sup>2</sup> No policy recommendation can get a top-grade rating unless it has top-ranked evidence claims in its favour. For these guides it is not only ‘ideal’ that  $P(e)$  be high if  $e$  is to count as evidence; it is necessary.<sup>3</sup>

<sup>1</sup> For instance evidence evaluation schemes from the Scottish Intercollegiate Guidelines Network, the International Agency for Cancer Research, or the Maryland rules in criminology.

<sup>2</sup> The guides clearly seem to make the assumption that high probability can be assigned to results from reliable methods. While an important issue, this is tangential to concerns here about whether a claim must have high probability if it is to be counted as evidence.

<sup>3</sup> This raises a question as to how to interpret probabilities. Roush claims that her definition of evidence is compatible with both subjectivist and objectivist readings. In Sect. 3 we explore issues relating to the subjectivist interpretation of probabilities when we consider Roush's argument from a Bayesian standpoint. To interpret Roush in an objectivist way, we avoid standard controversies in the philosophy of probability by assuming that  $e$  and  $h$  both denote event-types. This, however, is not to say we believe that it is straightforward to find an interpretation of probability that makes sense of Roush's definition of evidence or her arguments for probable evidence. Indeed, the difficulties discussed in Sect. 3.2 suggest otherwise.

There is a simple reason why one might adopt this view. It is almost certainly what motivates the ranking schemes and it is one of the reasons Roush herself gives: If  $e$  is to be good evidence for  $h$ ,  $e$  should provide good reason to believe  $h$ . Surely we shouldn't believe  $h$  on the basis of  $e$  unless there is good reason to believe  $e$ . So  $P(e)$  should be high. We shall here accept this line of defence for high  $P(e)$  without discussion and concentrate on the rest of Roush's discussion, for she has far more than this to offer. In particular she develops two original, challenging defences of high  $P(e)$ , both of which open new perspectives on the age-old topic of evidence. The first is based on an interesting mathematical relationship and a related series of graphs and the second on arguments against modelling surprising evidence as evidence with low probability.

We shall argue that these defences do not carry the conclusion. In good part that is because there is not one conclusion in Roush's discussion but three, all expressed in the same words: High  $P(e)$  is ideal if  $e$  is to be evidence for  $h$ . We claim that there are three conclusions because we think there are three different senses of 'evidence' at play in Roush's discussion, senses that are important to distinguish independent of their role in the specific issue of  $P(e)$ . They are:

- (1) Evidence as *the ground for knowledge*.<sup>4</sup> In order for  $e$  to be evidence for  $h$ ,  $e$  should be an appropriate basis for knowledge that  $h$ . The version of 'evidence as the ground for knowledge' we find in Roush supposes both that  $h$  be true—"as it must be for anyone to know it" (Roush 2005, p. 153) and also that  $e$  provides grounds for believing it so.<sup>5</sup>
- (2) Evidence as a two-place *relevance relation* (' $e$  is evidence for  $h$ ') between propositions or possible events, in which the evidence is supposed to be relevant to the truth of the hypothesis, without any presumption about whether either the evidence or the hypothesis is true.
- (3) Evidence for a hypothesis  $h$  as a *lever* to infer  $P(h)$ , that is, knowledge about the evidence or its probability can be used to deduce informative, previously unknown constraints on  $P(h)$ , or better,  $P(h)$  itself.

The second is the usual topic of confirmation theories and one could take it that Roush's explication is aimed here since she engages with the conventional literature at various points. It is at any rate an important topic, and again, not one just of philosophical interest. Consider hypothesis testing or policy deliberation. Gathering facts, finding out what is true and what is not, conducting experiments, even just sitting and discussing the facts is costly and time consuming. So one wants a concept of evidence

<sup>4</sup> The idea here is similar to Williamson's 'E = K' thesis (2000, p. 185) that evidence is just what we know. This thesis is motivated, as with Roush, by a desire to use evidence to justify belief in a hypothesis. Williamson's treatment is similar to Roush's in other respects as well. For instance, he requires that  $P(h|e) > P(h)$  for  $e$  to be evidence which is equivalent to Roush's discrimination condition. However, he does not adopt a condition similar to Roush's indication condition  $P(h|e) > 1/2$ . Instead Williamson uses the fact that evidence is knowledge and the requirement that  $P(h|e) > P(h)$  to justify belief in the hypothesis from evidence

<sup>5</sup> As Roush puts it in her discussion of tracking and evidence "... if  $h$  is true—as it must be for anyone to know it—and  $e$  tracks  $h$  then it is unlikely that  $e$  is false. And, if  $e$  is false, then because the subject's belief in  $e$  tracks  $e$ , the subject is unlikely to believe  $e$ . Since  $b(h)$  tracks  $b(e)$ , the probability of  $b(h)$  given  $b(e)$  is low too. All of this suggests that if the subject knows  $h$  through this trajectory, then because in order to do that she must believe  $h$ ,  $e$  is likely to be true." (Roush 2005, p. 153).

that tells what facts bear on the hypothesis in order to decide which ones to find out about, which experiments to run or which facts to let onto the table for discussion. This is looking at evidence from the perspective of the deliberation process, prior to any views about whether what is admitted as evidence provides sufficient grounds for believing the hypothesis, i.e. before considerations about issue (1). This perspective also fits particularly nicely with Roush's own concerns, which we separate out as issue (3), that evidence should provide leverage. She does not want  $P(h)$  to be presupposed in our attempts to settle if her two central requirements for evidence are met because that would undermine our ability to leverage from the evidence to the hypothesis.

Roush's discussion of high  $P(e)$  does not differentiate these three notions, yet  $P(e)$  seems to fare differently in each. For sense 1 it seems natural to suppose evidence should have high probability for the trivial reason that  $e$  can hardly be the basis for knowledge that  $h$  if  $e$  isn't itself true, or highly probable, just as the evidence-ranking schemes suppose. But high  $P(e)$  should surely be omitted as a criterion for evidence in sense 2. For sense 3, we shall argue, none of Roush's three criteria are necessary.

We look at Roush's defences of high  $P(e)$  in Sect. 3, evaluating them both on their own merit and with an eye to disentangling sense 1 and 2 of 'evidence'. We look at leveraging in Sect. 4. In Sect. 5 we produce a simple counterexample to show that, strictly speaking, evidence for  $h$  as defined by Roush does not imply that  $e$  tracks  $h$ , contrary to her hopes.

## 2 Roush's graphical analysis and definition of evidence

Roush constructs her definition of evidence from two desiderata. The first is that evidence should *discriminate* between hypotheses. She takes this to mean that if  $e$  is evidence for  $h$  then  $P(e|h) > P(e|\neg h)$ , or, in terms of the likelihood ratio ( $LR = P(e|h)/P(e|\neg h)$ ) that  $LR > 1$ .<sup>6</sup> Roush takes the discrimination condition to be uncontroversial and focuses greater attention on a second desideratum, the *indication* condition,  $P(h|e) > 0.5$ . The motivation for this is that evidence should, when true, make the hypothesis more likely than not, thus giving us some reason to believe  $h$  (rather than its alternative  $\neg h$ ).<sup>7</sup> These are both well rehearsed conditions, familiar from debates about how to define what we have prised out under the label 'evidence as a relevance relation'. For present purposes we shall take her arguments for them as sufficient for evidence in a relevance sense since our focus is on her further requirement that  $P(e)$  be high.

Roush defines evidence so that both desiderata are met, with particular emphasis on the indication condition.

*R*:  $e$  is some/good evidence for  $h$  if and only if "there is a lower bound greater than 1 on  $[LR]$  and a lower bound greater than 0 on  $P(e)$  such that  $P(h|e)$  is

<sup>6</sup> She further invokes a number of authors to argue that the likelihood ratio is the best measure of how good evidence is at discriminating.

<sup>7</sup> Roush notes "... we do not have good reason to believe, or even some reason to believe, a hypothesis is true, if we have no assurance that the posterior probability  $[P(h|e)] > 0.5$ " (Roush 2005, p. 165).

greater than 0.”/“greater than some high threshold appropriate to having good reason to believe” (Roush 2005, p. 183).

Despite its roundabout formulation in terms of lower bounds, the definition is logically equivalent to the following simpler definition.

(DEF1) *e* is *some*[alternatively, *good*] *evidence* for *h* if and only if

**DC** (*Discrimination Condition*):  $LR > 1$

**IC** (*Indication Condition*):  $P(h|e) > 0.5$  [alternatively,  $P(h|e) > a$ , where *a* is some chosen level greater than 0.5]

But this definition does not sit well with other statements made by Roush:

“An obvious solution ... is to adopt as a second condition for *e* to be evidence the demand that ...  $P(h|e)$ , be high. *However, that is merely a restatement of our desideratum.*” (emphasis added, Roush 2005, p. 166).

Also, the lack of reference to a lower bound on  $P(e)$  in (DEF1) ignores the importance Roush attributes to evidence being probable:

*R-addendum*: “high  $P(e)$  is not necessary but is ideal” (Roush 2005, p. 183).

This suggests that (DEF1) is not an adequate interpretation of Roush’s definition of evidence. An alternative, more suitable definition can be discerned by careful consideration of Roush’s graphical analysis, where she uses a series of graphs to explain the connection between lower bounds on  $LR$  and  $P(e)$  and **IC**. These are based on an identity that she establishes using the probability axioms:

$$A. \quad P(h|e) = [LR - P(e|h)/P(e)] / [LR - 1].$$

She points out that **A** implies facts about how  $P(h|e)$  can increase under special circumstances. The special circumstances are that

- (1)  $LR > 1$
- (2)  $LR$  is held fixed
- (3)  $P(e|h)$  is held fixed.

Note that this implies that  $P(e|\neg h)$  is also fixed.

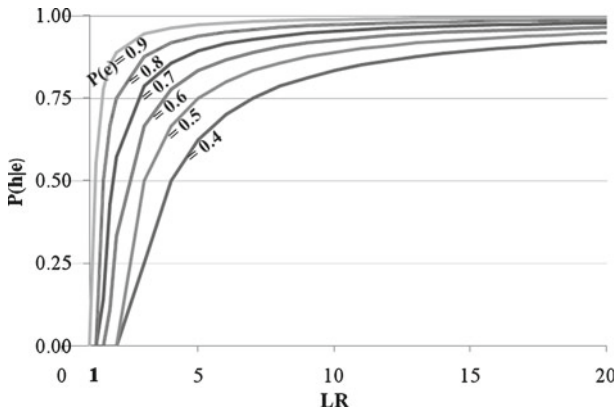
Given these three conditions it follows from **A** that

$$B. \quad P(h|e) \text{ increases with increasing } P(e).$$

Roush elaborates on **B** defending her advice that  $P(e)$  should ideally be high by taking the reader through a series of graphs and formulae.

We summarize her eight graphs in Fig. 1.

Her graph for  $P(e) = 0.4$  is the one she presents first. On it is displayed a shaded area above the  $P(e) = 0.4$  line in Fig. 1, representing the continuum of lines graphing  $P(h|e)$  versus  $LR$  for the continuum of values of  $P(e)$  possible above  $P(e) = 0.4$ .



**Fig. 1**  $P(h|e)$  versus LR for various fixed  $P(e)$  with  $P(e|h) = 1$

She explains:

“This graph presents a convenient lower limit for the trends that we will see when we increase LR and  $P(e)$ . The result is this: this surface bounds from below in the  $[P(h|e)]$  dimension every graph with  $LR > 1$  and  $P(e) > 0.4$ , and as these terms increase the  $[P(h|e)]$  term increases. That is, as LR and  $P(e)$  increase above 1 and 0.4, respectively, the value of  $P(h|e)$ , for any given values of  $P(e|h)$  and the LR, monotonically rises to 1. We can see why this is by inspecting the equation

$$P(h|e) = [LR - P(e|h)/P(e)] / (LR - 1)$$

If we suppose that the LR is some fixed value greater than 1, then  $P(e|h)$  will be greater than or equal to  $P(e)$ ... In other words, increasing  $P(e)$  with fixed or rising LR will have the effect of increasing  $P(h|e)$ .” (Roush 2005, p. 168)

This means that with fixed LR, high enough  $P(e)$  guarantees whatever value of  $P(h|e)$  is demanded. Specific values for  $P(h|e)$  are necessary on Roush’s characterization for  $e$  to be some/good evidence for  $h$ . Supposing  $LR > 1$ , it follows that high enough  $P(e)$  is a sufficient condition for a condition (viz.  $P(h|e) > a$ ) that is necessary for  $e$  to be some/good evidence for  $e$ . High  $P(e)$  is not necessary though, as Roush herself notes in places. We stress this because we found some comments in her text that could be misleading on this.<sup>8</sup>

Roush’s graphical analysis also suggests a corresponding definition of evidence. In her analysis, the lower bounds on LR and  $P(e)$  that are sufficient for **IC** are independent of the value of  $P(e|h)$ . Indeed, she states as much: “Through graphing  $P(e|h)$ ,  $P(e)$  and the LR, we have found a result that is independent of  $P(e|h)$  and depends only on  $P(e)$  and LR.” (Roush 2005, p. 168) Though she does not say this explicitly,

<sup>8</sup> For instance, when she says “my proposal, then, is that the second condition on evidence, the indication condition, be a lower bound on the value of  $P(e)$ ”. Or just after, “there are three broad questions to ask about this idea... a third is whether high  $P(e)$  is plausibly a necessary condition for evidence, since there seem to be counterexamples” (Roush 2005, p. 171).

this suggests that Roush intends the lower bounds on LR and P(e) in her definition of evidence to be independent of the value of P(e|h). To construct a corresponding formal definition of evidence, it is first instructive to construct lower-bound conditions equivalent to **DC** and **IC**. In the appendix we prove:

C. **DC** and **IC** if and only if there exist  $x > 1, y > 0$  and  $0 < z \leq 1$  such that

- (i)  $LR \geq x$
- (ii)  $P(e) > y$
- (iii)  $P(e|h) \leq z$
- (iv)  $\frac{x - z/y}{x - 1} = 1/2$

Condition (iii) shows how, given lower bounds on P(e) and LR, whether **DC** and **IC** are met depends on P(e|h). Roush’s apparent desire to construct a definition of evidence in terms of lower bounds that are independent of P(e|h) suggests the following natural move: To stipulate that  $z=1$  so that condition (iii) becomes  $P(e|h) \leq 1$ , which always holds regardless of the value of P(e|h). This suggests defining evidence this way:<sup>9</sup>

(DEF 2) e is *some* [alternatively, *good*] evidence for h if and only if there exist  $x > 1, y > 0$  such that

- (i)  $LR \geq x$
- (ii)  $P(e) > y$
- (iii)  $\frac{x - 1/y}{x - 1} = 1/2$  [alternatively  $\frac{x - 1/y}{x - 1} = a$  for appropriate  $a > 1/2$ ]

(DEF2) fits Roush’s roundabout expression and gives P(e|h)-independent lower bounds in line with her graphical analysis. It also makes explicit a trade-off: As the lower bound on LR strengthens, the lower bound on P(e) can weaken and *vice versa*. Given this trade-off, high P(e) is not necessary for e to be evidence, since for any low y there is a sufficiently high x that ensures that e is evidence. Yet for any given x, a higher y raises the lower bound on P(h|e) setting out how more probable evidence is ideal. In these ways the above definition of evidence neatly fits Roush’s discussion.

In this definition **IC** and **DC** are necessary but not sufficient for evidence as can be seen in the following probability distribution:

$$P(e) = 0.6001, \quad P(h) = 0.8, \quad LR = 2$$

For this distribution  $P(h|e) \approx 8/9$ . But e is not evidence because the lower bound on P(e) would have to be less than 0.6001 and by (iii) this implies that the lower bound on LR would have to be greater than  $2^{1/3}$ , which is false. This example shows that this version of Roush’s definition has an undesirable consequence of ruling out some cases where e is probable and **IC** and **DC** are met. Why shouldn’t e count as evidence in cases like this?

What is missing here is why a definition of evidence should be constructed in terms of lower bounds of LR and P(e) that are *independent* of the value of P(e|h). Though Roush does not explicitly discuss this, one candidate answer is that it yields a leverage

<sup>9</sup> This reading is in accord with suggestions from an anonymous referee.

advantage by it allowing one to classify  $e$  as evidence without needing to know  $P(e|h)$ . Roush's discussion of the leverage advantage of using a lower bound on  $P(e)$  (Roush 2005, p. 170) suggests that this may be her motivation. However, this leverage advantage comes at a cost, since it rules out the case above, which seems paradigmatically evidence on Roush's terms (*IC* and *DC* are met,  $P(e)$  is high). Moreover ignorance of  $P(e|h)$  also makes it hard to evaluate LR!

Finally, it should be stressed that, though Roush's definition requires the existence of a lower bound on  $P(e)$ , it does not require that evidence be probable at all, since the lower bound on  $P(e)$  can—provided LR is high enough—be arbitrarily close to zero. So, although probable evidence is defended by Roush as ideal, it is not a necessary condition of evidence as she defines it. We think this is an advantage, since we now defend improbable evidence.

### 3 In defence of improbable evidence

We find in Roush three major arguments that high  $P(e)$  is ideal:

- The argument from equation *A* and the accompanying graphs.
- An argument against a simple story of Bayesian updating.
- High  $P(e)$  has a leveraging advantage for finding  $P(h)$ .

The first argument seems most suited to a relevance sense of evidence, but we discuss it in Sect. 3.1 more or less on its own grounds without putting weight on our view that there are three different senses of evidence involved in Roush's discussion. The second argument seems geared to knowledge as a grounds for knowledge. Bayesian updating allows that surprising evidence—in the sense of evidence with low probability—can increase the degree of belief in a hypothesis more than non-surprising evidence. Roush's discussion seems to suppose that this is incompatible with her view that evidence should have high probability. We address this in Sect. 3.2. Section 3.3 briefly considers other shorter defences Roush offers for high  $P(e)$ . We take up leveraging in Sect. 4.

#### 3.1 Low probability evidence can satisfy *DC* and *IC* maximally

We assume, for the sake of argument, that high  $P(e)$  is a reasonable requirement on evidence as a ground to knowledge. But we do not see how it can be taken as an ideal way to satisfy the definition Roush offers, which at any rate we think is best seen as a reasonable candidate for defining a relevance notion of evidence. For despite Roush's proposal, a lower probability claim can make for better evidence using her own criteria.

Suppose  $P(e|h) = 1$ ,<sup>10</sup> which is one way to model 'h explains e' in the deductive-nomological account of explanation. Then Bayes Theorem reduces to

<sup>10</sup> Similar examples can be generated for any fixed non-zero value of  $P(e|h)$ . Note that given  $P(e|h) = 1$ ,  $e$  is evidence if and only if it satisfies *DC* and *IC*.



$$D. \quad P(h|e) = P(h)/P(e).$$

Since  $P(e|h) = 1$ , it follows that

$$E. \quad P(e) = P(h) + P(e|\neg h)P(\neg h)$$

and so

$$F. \quad P(h|e) = P(h) / [P(h) + P(e|\neg h)P(\neg h)].$$

Given  $P(e|h) = 1$  it also follows that

$$G. \quad LR = 1/P(e|\neg h).$$

So lowering  $P(e)$  by lowering  $P(e|\neg h)$  simultaneously produces improvements in LR and in  $P(h|e)$ , making  $e$  better evidence for  $h$  on both Roush's criteria. While it is true, as she concludes, that "increasing  $P(e)$  with fixed or rising LR will have the effect of increasing  $P(h|e)$ " (Roush 2005, p. 168), it is equally true that decreasing  $P(e)$  with rising LR can have the effect of increasing  $P(h|e)$ . Thus, the graphs hardly provide a strong argument for increasing  $P(e)$  in order to satisfy the criteria for evidence.

Not only can lowering  $P(e)$  raise both LR and  $P(h|e)$ , but both conditions **DC** and **IC** can be maximally satisfied while  $P(e)$  takes any value whatsoever. Suppose  $e$  is a perfect sign of  $h$ ; i.e.  $e \equiv h$ . Then  $P(h|e) = 1$  and LR is infinitely high, but  $P(e)$  can be as small or as large as one would like. This example has another nice aspect. Whenever there are two independent criteria for the same thing, trade-offs may be required, but here a trade-off is avoided. In this case (or any case with fixed  $P(e|h)$ ) less probable evidence can be better evidence by both criteria at once.

We should also note that Roush's graphical arguments for high  $P(e)$  depend on the asymmetry with which she treats the two independent criteria for evidence. Suppose  $e$  is 'candidate' evidence for  $h$  in the sense that **DC** is well satisfied (i.e. LR is high). Then high  $P(e)$  is sufficient for the satisfaction of **IC**. But the exactly symmetric claim is not true. Suppose  $e$  is 'candidate' evidence for  $h$  in the sense that **IC** is well satisfied. Then it is not true that high  $P(e)$  is sufficient for the satisfaction of **DC**.<sup>11</sup> So high  $P(e)$  is useful for obtaining high  $P(h|e)$  when LR is sufficiently high, but high  $P(e)$  is not sufficient for high LR when  $P(h|e)$  is high. Yet there seems to be no special reason for considering either criterion differently from the other.<sup>12</sup>

### 3.2 A Bayesian defence of improbable evidence and a frequency defence of probable evidence

As a prelude to our arguments, we first set out two distinct analyses on how the probability of evidence relates to the probability of the hypothesis. The first is Roush's while the second is an analysis often presented to support the conventional claim that evidence with a lower probability makes a hypothesis more probable (all else being equal) than evidence with a higher probability.

<sup>11</sup> See Theorem 2 in the appendix.

<sup>12</sup> We have not here rehearsed Roush's arguments for **DC** and **IC** but we don't find anything in them that gives a reason to treat the two in this different way.

**Table 1** Two analyses of the relationship between P(e) and P(h|e)

Analysis	Fixed factors	Change to P(e)	Resulting change in P(h e)	Resulting changes in other factors
Roush	LR (>1), P(e h)	P(e) increases	P(h e) increases	P(h) increases, P(e ¬h) fixed
Conventional	P(e h), P(h)	P(e) decreases	P(h e) increases	LR increases, P(e ¬h) decreases

Section 2 described Roush's argument that if LR is sufficiently high a lower bound on P(e) is sufficient for a lower bound on P(h|e). Importantly, given Roush's constraints—P(e|h) fixed and LR fixed (>1)—P(e) increases if and only if P(h) increases.

A second analysis, one conventionally used in discussions of the greater confirmation power of surprising evidence,<sup>13</sup> follows from Bayes Theorem:

$$P(h|e) = P(h)P(e|h)/P(e)$$

In this case, assuming P(e|h) and P(h) are fixed, P(h|e) increases as P(e) decreases. In this case, LR must increase and P(e|¬h) decrease when P(e) decreases. Labelling evidence that has lower probability as 'more surprising', this result shows that the more surprising e is, the higher P(h|e) is.

These two analyses can be summarised in Table 1.

Since both analyses follow from the probability axioms, there is no contradiction between them despite their apparently conflicting conclusions as to the relationship between changes in P(e) and changes in P(h|e). As the table makes clear, the difference is due to different factors being held fixed.

This is just arithmetic with probabilities. However, both analyses are used to make arguments as to the significance of the probability of evidence. The conventional analysis is used to argue, via Bayesian updating, that the more surprising evidence is, the more confirmation it lends to the hypothesis once learned. This clearly involves evidence in the sense of grounds to knowledge. Roush's is used to support her claims that more probable evidence is 'ideal' and if there is to be a conflict at all, this must involve evidence in the same sense. To clarify the dispute we shall first consider what Roush's analysis looks like in a 'Bayesian' framework, then what it looks like in a 'frequentist' framework. By a 'Bayesian framework' we mean one in which probabilities represent degrees of belief and in which on learning a new fact e, probabilities are 'updated' by changing from the 'prior' probability (labelled P<sub>i</sub>(●)) to a new 'posterior' probability (labelled P<sub>f</sub>(●)) by the rule : P<sub>f</sub>(●) = P<sub>i</sub>(●|e).

Interpreted in a Bayesian way, Roush's analysis can then be expressed as follows. If the agent holds a higher prior degree of belief in e, but the same values for P<sub>i</sub>(e|h) and P<sub>i</sub>(e|¬h), then on learning e the agent would have a higher posterior degree of belief in h than would have been the case had the agent learned e while holding the lower prior belief in e. Though this is a consistent Bayesian account of how higher priors in e can be advantageous for obtaining a higher posterior in h on learning e,

<sup>13</sup> See, for example, (Howson and Urbach 2005, p. 97).

there is problem. The agent can only have the higher prior in  $e$ , given the other fixed conditional probabilities, if the agent also has a higher prior in the hypothesis. This follows from the fact that the higher  $P_i(e)$  also implies a higher  $P_i(h)$  given the factors Roush holds fixed. There is then no reason to attribute the higher posterior in the hypothesis to the higher prior in the evidence (as Roush would like) rather than to the higher prior  $P_i(h)$ .

In contrast the conventional analysis avoids this difficulty. Here  $P_i(h)$  is fixed when comparing the cases where evidence has low and high prior degrees of belief. Thus, the higher posterior degree of belief in  $h$  once  $e$  has been learned *is* attributable to the lower prior degree of belief in the evidence under the conventional view. Indeed, this is just the Bayesian story as to why surprising evidence confirms more: Evidence with a lower prior once learned raises the posterior in the hypothesis more assuming fixed values for the probability for the hypothesis and for the conditional probability of the evidence given the hypothesis.

Unsurprisingly perhaps, given this tension between her analysis and a Bayesian interpretation, Roush rejects the premise that  $P(h)$  should be fixed when comparing high and low probability evidence. Instead she takes it that high  $P(e)$  should make a difference to  $P(h)$  *before* learning that  $e$  is the case:

“Your degree of belief in  $e$  prior to the conditionalisation is just  $P(e)$ , so high  $P(e)$  is (almost) sufficient for you to take  $e$  as evidence for whatever  $e$  happens to be positively relevant to, that is, to conditionalize upon it. Roughly, if you are confident of  $e$ , then you ought to let your other beliefs feel the appropriate effects of  $e$ 's truth” (Roush 2005, p. 174)

However, for a Bayesian this is irrational, since if  $P(e)$  is less than one then this *means* that the agent does not believe  $e$  is certain and would not rationally “let [their] other beliefs feel the appropriate effects of  $e$ 's truth”.<sup>14</sup>

Roush, however, takes her unorthodox interpretation of Bayesian updating to be virtuous, since it fits with her view that for  $e$  to be evidence for anything else it must itself be probable:

“It seems to me inescapable that in order for the value of  $P(e)$  that precedes Bayesian strict conditionalization to justify Bayesian strict conditionalization  $P(e)$  must be high” (Roush 2005, p. 174).

It is as if Roush supposes that Bayesians have a three-step process. Agents begin with degrees of belief represented by the ‘antecedent’ probability  $P_i$ . At the first step they observe  $e$ . At the second they decide on this basis that the probability of  $e$  should be 1. Because the probability of  $e$  is 1 they are justified, at the third step, in changing their degrees of belief to those represented by the ‘posterior’ probability  $P_f$ . But of course Bayesians do not take three steps, only two. They observe  $e$  at the first step and at the second, revise their probabilities in one fell swoop to  $P_f$ , which among other features sets the probability of  $e$  to 1. For the Bayesian the new probability is justified

<sup>14</sup> Note that Jeffrey conditionalization does not help here, since conditionalization on the original value of  $P(e)$  does not lead to any change in the degrees of belief, and conditionalizing on a different value of  $P(e)$  is inconsistent with the agent's degree of belief in  $e$ .

by learning  $e$ , not by the fact that one has become confident of  $e$  (i.e., already set the probability of  $e$  high). The posterior probability is an expression of one's confidence, not a justification of it. The Bayesian is far more objective here than Roush would have it: It is observations that justify new degrees of belief, not simply one's antecedent degrees of confidence.

These difficulties suggest a possible fix: Do not assume  $e$  is certain when it is not, but use Jeffrey conditionalization, under which  $P_f(\bullet) = P_i(\bullet|e)P_f(e) + P_i(\bullet|\neg e)P_f(\neg e)$ . With this approach one might be able to argue that it is better to have a higher prior in the evidence, assuming identical  $P_i(e|h)$  and  $P_i(e|\neg h)$ , than otherwise. However, as it happens this runs into the same difficulty as the strict conditionalization case, namely a higher prior in the hypothesis is implied when there is a higher prior in the evidence. Thus, even with Jeffrey conditionalization, the higher posterior cannot be attributed to the higher prior in the evidence.<sup>15</sup>

Perhaps a resolution lies elsewhere. In her examples<sup>16</sup> Roush describes scientists or doctors finding out that the evidence is probable, arguing that having a high probability here is important. This may suggest that the way to make sense of the importance of probable evidence for Roush is take evidence to be probable *once the agent has become confident of it*, that is, to take her demand that the probability of the evidence to be high to refer to the posterior rather than prior probability of the evidence.<sup>17</sup> This would assume Jeffrey conditionalization, since in the strict updating case the posterior of the evidence is always one, so more or less probable evidence cannot be modelled.<sup>18</sup>

At first blush, this modified approach looks promising. To see why, assume identical priors in the evidence and in the hypothesis in order to allow one to attribute the greater confirmation power to the higher probability of evidence. With Jeffrey conditionalization, the higher the posterior in the evidence, the higher the posterior in the hypothesis.<sup>19</sup> Yet this is not consistent with Roush's analysis, since under these assumed conditions,  $P_i(h|e)$  (which under Jeffrey conditionalization equals  $P_f(h|e)$ ) must be the same across the comparison of higher and lower probability evidence. Moreover, this proposed analysis amounts to an argument that it is better to learn more probable evidence because it raises the posterior in the hypothesis more. But since  $P(h|e)$  is unchanged, this is not an argument that probable evidence makes for better evidence in a relevance sense. It is rather an argument that learning more probable

<sup>15</sup> This is as one would expect given that it is a generalisation of the strict Bayesian case already discussed.

<sup>16</sup> Such as the Rutherford example (Roush 2005, p. 174) and her hypothetical medical test example (Roush 2005, p. 171).

<sup>17</sup> Such a reading also fits well with some of Roush's comments: "P(e) reports actual degree of belief, not how much you expected at some prior stage that you would believe  $e$  at this stage" (Roush 2005, p. 175).

<sup>18</sup> Yet another option would be to move to Jeffrey conditionalization, take  $P(e)$  to be a posterior but stick to something akin to Roush's three-step updating by allowing the updating to  $P(e)$  before updating other degrees of belief on  $e$ . This might better describe how beliefs change in practice and could also help modelling 'old evidence' situations, since old evidence could be modelled as evidence which has already been updated upon specifically, but which has not been updated upon generally. However, this approach also requires that the agent hold incoherent degrees of belief which undesirable when one is arguing for a normative model of belief, as Roush is doing.

<sup>19</sup> This follows immediately from the formula for Jeffrey conditionalization: i.e. the greater  $P_f(e)$  is (assuming  $LR_i > 1$ ), the greater  $P_f(h)$  will be.

evidence makes for better grounds-for-knowledge evidence since it leads to a higher posterior in  $h$ . But that more probable evidence makes for better grounds for knowledge is not in dispute.<sup>20</sup>

Turn finally to the frequency perspective on probability, from which Roush's argument for probable evidence *can* be made sense of. Consider two populations where event-types  $h$  and  $e$  both occur, where  $P(e|h)$  and  $P(e|\neg h)$  are the same across both populations, and where  $LR > 1$ . It follows by Roush's analysis that if  $e$  is more long-run frequent—in this sense more probable—in the first population, then the probability that  $h$  occurs in the subpopulation where  $e$  occurs must also be greater for that population. This shows that if an event-type is more probable (in the frequency sense), it must also be more positively relevant to  $h$ , assuming  $P(e|h)$  and  $P(e|\neg h)$  are the same (again in the frequency sense).

Note that by contrast with the Bayesian cases above, the fact that  $h$  must also be more probable in the population where  $e$  is more probable is not a problem. Far from it,  $h$ 's being more probable in the first population shows that, in addition to more probable evidence making for better relevance evidence (in that  $P(h|e)$  is higher), it makes for better grounds for knowledge in that  $P(h)$  is higher. We must be careful, however, about what claim ' $h$ ' represents. That  $P(h)$  is higher in one population than another gives better grounds for knowledge of the claim that a randomly drawn member of that population will be an  $h$ . Likewise the other probabilities,  $P(e)$  and conditional probabilities, must be interpreted accordingly for the same population.<sup>21</sup>

Finally, it is important to emphasise that not all of Roush's examples of hypotheses can be construed as event-types occurring in (ideally) 'infinite' populations. Nor can most hypotheses for which we wish to have a theory of evidence. There are notorious and well-rehearsed difficulties in applying this kind of frequentist account to hypotheses of arbitrary form, which we need not repeat here.

### 3.3 What we conclude about Roush's defences of high $P(e)$

In conclusion, independently of her arguments about tracking knowledge (which we turn to in Sect. 5), Roush defends her claim that  $P(e)$  be high in order for  $e$  to be evidence for anything on three fronts. The first involves arguments based on formula **A**. and the accompanying graphs. These, we have argued, provide weak grounds for the demand, if any at all. On the second front she attempts to defuse arguments to the opposite conclusion, that  $P(e)$  should be low. One of her arguments on this front is that

<sup>20</sup> A final attempt to find a suitable Bayesian interpretation of Roush's argument might be to try to show that a higher posterior in the evidence after Jeffrey conditionalization makes for higher  $P_f(h|e)$  assuming identical posteriors for  $P_f(e|h)$ ,  $P_f(e|\neg h)$  such that  $LR_f > 1$  in line with Roush's formal analysis. We then know, from Roush, that  $P_f(h|e)$  must be greater when  $P_f(e)$  is greater. But this also fails to give a plausible Bayesian argument for probable evidence since it implies a higher prior in the hypothesis for the case where one learns the more probable evidence (see Theorem 4 in appendix).

<sup>21</sup> The importance of taking care when interpreting the probabilities can be seen in Eric Barnes' recent criticism of Roush's medical example (Roush 2005, p. 171) where he argues that she equivocates in her interpretation of  $P(e)$ . See (Barnes 2008, pp. 554–555) for details.

Bayesians need high  $P(e)$  to warrant updating degrees of belief. We have countered that this argument rests on a mistake about the nature of Bayesian conditionalization. In a further attempt to reconcile Roush's arguments for probable evidence with Bayesianism, we also considered Jeffrey conditionalization. The first case considered, Jeffrey conditionalization with a higher prior in the evidence, entailed a higher prior in the hypothesis undermining the attribution the greater confirmation power to the higher probability of the evidence. The second case, Jeffrey conditionalization with a higher posterior in the evidence, entailed a higher posterior in the hypothesis with  $P(h|e)$  unchanged, thus showing how learning more probable evidence can be beneficial in a grounds-for-knowledge rather than relevance sense.

In addition to these two fronts, Roush points out that with  $LR > 1$  as the criterion, as opposed to  $P(h|e) > P(h)$ ,  $e$  can still discriminate even if it has probability very close to 1. She also offers an alternative interpretation to some examples of Peter Achinstein that were supposed to provide cases where "it is the very improbability of  $e$  that makes it evidence for  $h$ " (p. 176). All these show either that  $P(e)$  need not be small or that it is no harm for it to be big. This is in line with the view that follows from *DC* and *IC*, that the probability of  $e$  is irrelevant to whether  $e$  is evidence or not.

The only positive argument for evidence—in a relevance sense—to be probable was revealed when we considered a frequentist interpretation of Roush's argument. Here it was shown that higher probability evidence can be beneficial both in a grounds-for-knowledge and relevance sense. This positive result, however, is somewhat tempered by the well known limitations of frequentism for assigning probabilities to hypotheses.

#### 4 Leveraging

We want to use evidence for  $h$  to arrive at an assessment of  $P(h)$ —that's what evidence is supposed to be good for. If the ultimate aim is to use the evidence to arrive at an assessment of  $P(h)$ , it should not be necessary to assign a value to  $P(h)$  in order either to assess  $P(e)$  or to assess whether  $e$  is evidence for  $h$ . In Roush's terminology, we should *leverage* to  $P(h)$  by using information other than  $P(h)$ .

For Roush, evidence must meet both *DC* and *IC*. To assess if *DC* is met, it is sufficient to assess  $LR$  and the usual and most immediate way to do this is to assess  $P(e|h)$  and  $P(e|\neg h)$ . Formula *A* shows that all that is required in addition to assess whether *IC* is met is an assessment of  $P(e)$ . And the discussion following it shows that high enough  $P(e)$  ensures *IC*, given that *DC* is met. So an ideal way to satisfy *DC* and *IC* is for  $P(e)$  to be high. If we know this, we can know that  $e$  is evidence for  $h$  without having to assess  $P(h)$ .

Leveraging is an idea we entirely endorse (cf. Cartwright (2009)). Indeed the importance of leveraging cannot be stressed enough when it comes to considerations of the *use* of evidence, considerations that we think philosophers need to keep centrally in view in developing accounts of what evidence 'really is'. Both pure science and policy want to use evidence for  $h$  to help to arrive at a reliable estimate of  $P(h)$ . This gives yet another argument, based on the idea of leveraging, for concentrating as Roush does on the size of  $P(e)$ .

Suppose one has gone down the route of demanding that evidence must have a high likelihood ratio, as she supposes, or as many others suppose, must satisfy a relevance requirement, like  $P(h|e) > P(h|\neg e)$ . In both cases, once it is known that either of these requirements is met by knowing the relevant conditional probabilities, it remains only to learn  $P(e)$  to fix  $P(h)$  because both the following formulae are true:

$$P(h) = P(h|e)P(e) + P(h|\neg e)(1 - P(e))$$

$$P(h) = \frac{P(e) - P(e|\neg h)}{P(e|h) - P(e|\neg h)}$$

So from the point of view of leveraging, if one must know  $P(e)$  in order for  $e$  to be usable as evidence (as one must given Roush's requirement that  $P(e)$  is high), demanding knowledge either of the components of the likelihood ratio or the components of the relevance difference is enough. No additional requirements are needed, such as *DC*. More may be needed to characterize what it is for  $e$  to be 'relevant to' the truth of  $h$ , that is, for our second sense of evidence. But they are not needed for leveraging, which can make do with far less. This is why we think it is important to prise the two notions apart and to allow different accounts for them. In particular, for leveraging, tacking on requirements from the other senses of evidence can be highly restrictive. When it comes to calculating a given target probability, any kind of information that does the job is as good as any other. It all depends on what probability assignments we already accept or can efficiently find out. The calculus of probability constrains the relations among probabilistic facts, but a large variety of combinations can fix the value for a given target. This suggests that no particular constraints should be put on what probabilistic facts should be counted as evidence when it is evidence as leverage to targeted probabilities that is at stake.<sup>22</sup>

Leveraging has two aspects, of which we have so far discussed only one. For evidence of  $h$  to be of genuine use, not only should it help us calculate  $P(h)$ , but it should also be more accessible than  $P(h)$  itself. From this point of view we should like to comment, albeit briefly, on a deep and controversial position that Roush defends: that evidence should be characterized entirely in terms of probability. Roush argues that this should be done in order to avoid introducing concepts in the explication of 'evidence' that are even more obscure than 'evidence' itself. Indeed, she claims, concepts are often offered in explication of 'evidence' themselves generally receive their clearest explication in terms of probabilities. Explanatory relevance is a prime example.

We take issue with this last claim, since it has been argued at length that causation, and thereby causal explanation and thereby explanation in general, cannot be given a purely probabilistic explication.<sup>23</sup> But that is not the issue we would like to point to here. Rather we worry about the fact that probabilities are hard to come by. It is for this reason that we urge that the project of explicating 'evidence' should start a big step before the starting position of Roush and others who offer purely probabilistic accounts. For the purposes of both pure science and policy it is standard practice—and

<sup>22</sup> Note too that in this case it is not knowledge of events that is being employed but rather knowledge of their probabilities.

<sup>23</sup> Cf., among works by many authors, Cartwright (1979, 1989)



a practice we would wish to defend—to first gather the evidence and then to use it to assess various probabilities, e.g.  $P(h)$ ,  $P(h|e)$ ,  $P(h|\neg e)$ ,  $P(e|h)$  or  $P(e|\neg h)$ . And for this we need guidelines about what counts as evidence that are not couched in terms of probabilities. One might think of these guidelines as an approach to our second notion of evidence as a two-place relevance relation. But it also helps with the accessibility aspect of leveraging, since the idea is to isolate those kinds of facts that will help in the assessment of the otherwise difficult to reach probabilities. This, however, is a project much in its infancy in the contemporary philosophical literature.

## 5 Evidence and tracking

What is the connection between tracking and evidence as Roush defines it, and what role does high  $P(e)$  play in it? Very roughly,  $x$  tracks  $y$  means that  $x$  and  $y$  are correlated: They both obtain or fail to obtain together. Roush is concerned with the kinds of cases common in philosophy of science in which a subject comes to know  $h$  via believing evidence  $e$ . For this she maintains that high  $P(e)$  is required because

“...in this trajectory for knowing  $h$  not only is  $h$  true but also  $b(h)$  TRACKS  $b(e)$ ,  $b(e)$  TRACKS  $e$ , and  $e$  TRACKS  $h$ . Now, if  $h$  is true—as it must be for anyone to know it—and  $e$  TRACKS  $h$  then it is unlikely that  $e$  is false.” (Roush 2005, p. 153) [ $b(x)$  = ‘The agent believes  $x$ .’]

This involves claims about beliefs whereas the relationship between evidence and hypothesis does not involve belief. As one would expect therefore, the relevant concept of tracking for evidence is different. Following (Roush 2005, p. 150) it can be formulated as follows

Evidence  $e$  tracks a hypothesis  $h$  at level  $u$  ( $< 1$ ) if and only if

**TR1.**  $P(e|h) > u$

**TR2.**  $P(e|\neg h) < 1 - u$ .

For Roush, it is important that evidence tracks the hypothesis because that ensures—provided  $u$  is high enough and other tracking relations are met—the desirable epistemic goal that belief in the hypothesis tracks the hypothesis.<sup>24</sup>

Note that one could take **TR1** and **TR2** to define evidence and call it ‘tracking evidence’. By definition then, evidence would track the hypothesis and it would, following (Roush 2005, p. 154), meet intuitive indication and discrimination conditions. Tracking evidence is also an example of a relevance concept of evidence. Here whether or not  $e$  is tracking evidence for  $h$  depends purely on the relationship between  $e$  and  $h$ , and the probabilities of  $e$  and  $h$  are no part of the characterization. Moreover, it has the nice property that when  $e$  is tracking evidence, then  $P(e)$  and  $P(h)$  must be close in value. But, whether  $P(e)$  and  $P(h)$  are high or low is independent of whether  $e$  is tracking evidence for  $h$ .

<sup>24</sup> This follows from the ‘transitivity enough’ property of the tracking relation (Roush 2005, pp. 151–152).



Given the attractive properties of the tracking definition, one may wonder why Roush does not adopt it. In short, (Roush 2005, p. 160) explains that she does not define evidence in this tracking way because so defined,  $e$  being evidence for  $h$  does not imply  $P(h|e) > 0.5$ . Thus,  $e$  can be evidence for  $h$  and yet fail to provide adequate reason to believe  $h$ . Nevertheless Roush does not relinquish the aim that evidence should track the hypothesis. So it is important to ask whether evidence as she defines it meets the tracking requirement. It is not hard to see that it does not always do so. The probability distribution specified by  $P(e) = 0.2$ ,  $P(h) = 0.9$ ,  $LR = 19$  has  $P(h|e) = 0.994$ ,  $P(e|h) = 0.22$  and  $P(e|\neg h) = 0.012$ . It is evidence under Roush's definition, but given the low value of  $P(e|h)$ ,  $e$  does not track  $h$ .

More generally, the relationship between Roush's definition of evidence and tracking can be made clearer using two simple bounds on  $P(e|h)$  and  $P(e|\neg h)$ .<sup>25</sup>

- (i)  $1/LR \geq P(e|\neg h)$
- (ii)  $P(e)/P(h) \geq P(e|h) > P(e)$

When  $LR$  is sufficiently high, bound (i) implies that  $P(e|\neg h)$  must be low and thus that **TR2** will be met. Likewise, bound (ii) shows that a high  $P(e)$  is sufficient for **TR1**. This fits Roush's analysis well, since high  $LR$  and  $P(e)$  are shown together to be sufficient for evidence to track. Moreover, the higher  $LR$  and  $P(e)$  are, the better the tracking will be. However, bound (ii) also shows how tracking can fail when  $P(e)$  is significantly smaller than  $P(h)$  since then  $P(e|h)$  must be small so **TR2** fails. This is what happens in the example above.

The relationship between Roush's evidence and tracking suggests another rationale for imposing a lower bound on  $P(e)$ : To ensure that evidence tracks the hypothesis. However, doing this does not just imply tracking. To see why, recall that tracking evidence is a relevance concept. However, when supplemented with a requirement that  $P(e)$  be high, one can infer that  $P(h)$  must also be (quite) high in virtue of  $e$  tracking  $h$ . Therefore, imposing a high lower bound on  $P(e)$  ensures  $e$  tracks  $h$  and thus that  $P(h)$  is high,<sup>26</sup> which is what is required for a grounds-for-knowledge concept. So high probability of evidence plays a double role, which arguably leads to a conflation absent in the simple tracking concept of evidence. Probable evidence makes evidence track the hypothesis, a feature characteristic of evidence in the relevance sense, and simultaneously makes evidence a ground for knowledge.

To finish, it is interesting to note that Roush's failure to define evidence so that it implies that evidence tracks a hypothesis need not be a serious problem for her concept of evidence. Tracking evidence is extremely powerful when one has it, since if one knows the evidence is false, then one can be pretty sure the hypothesis is false, and conversely. This, though highly desirable, is rarely met in practice. Often evidence speaks for the truth of a hypothesis when we know it to hold, but when false does not say much for the falsity for the hypothesis. For example, let

<sup>25</sup> The bounds are derived as follows. First,  $LR = P(e|h)/P(e|\neg h)$ , so  $P(e|\neg h) = P(e|h)/LR$ . But since  $P(e|h) \leq 1$ , it follows that  $P(e|\neg h) \leq 1/LR$ . Second,  $P(e|h) = P(e \& h)/P(h)$ , but  $P(e \& h) \leq P(e)$  because  $e \& h \Rightarrow e$ , so  $P(e|h) \leq P(e)/P(h)$ . Finally because  $LR > 1$ , it follows that  $P(e|h) > P(e)$ .

<sup>26</sup> When  $LR > 1$ ,  $1 \geq P(e|h) - P(e|\neg h) > 0$ . But given  $P(h) = [P(e) - P(e|\neg h)] / [P(e|h) - P(e|\neg h)]$  it follows that  $P(h) \geq P(e) - P(e|\neg h)$ . Since high  $P(e)$  ensures tracking,  $P(e|\neg h)$  is low and thus  $P(h)$  must be at least almost as high as  $P(e)$ .

- h: Jill murdered Jack  
 e: Jill's fingerprints are on the murder weapon.

In this case, e is intuitively evidence for h. However, e does not track h: Suppose that Jill is a careful, intelligent person and if she had decided to murder someone she would have used gloves, so  $P(e|h)$  is low. Given a plausible probability distribution assignment to e and h here, e would be good evidence for h. So Roush's concept of evidence models this situation well. In contrast, the tracking concept is overly strong and rules the fingerprints out as evidence.

## 6 Conclusion

In this paper we have clarified Roush's definition of evidence and critically analyzed her arguments that evidence should be probable. Roush's first argument, based on formula A and the associated graphs, we think is weak. These show that high  $P(e)$  is sufficient but not necessary for e to be evidence. As we have shown, both of Roush's criteria for evidence can be met maximally and  $P(e)$  take any value at all. Our attempts to reconcile Roush's arguments for probable evidence with Bayesianism at best show that probable evidence makes for better grounds-for-knowledge evidence. In our analysis only a frequentist interpretation supports an argument that more probable evidence makes for better relevance evidence, but this is limited to cases where frequentist probabilities can be applied.

The latter part of our paper discussed the role of leveraging and the relationship between evidence and tracking. We strongly agree with Roush on the importance of leveraging, particularly when using evidence in policy. We have argued, however, that Roush's two criteria for evidence are unduly restrictive from a leveraging view. Almost any constraint on  $P(e)$  imposes a constraint on  $P(h)$  and thus serves the purpose of leveraging  $P(h)$ . As regards tracking we showed that evidence as Roush defines it can fail to track the hypothesis. If Roush's definition is supplemented with a requirement that evidence be probable then evidence does track the hypothesis. This suggests another reason why Roush may see probable evidence as ideal. However, since the concept of tracking evidence is restrictive this provides weak grounds at best for taking probable evidence as ideal.

In sum it may be the case that high probability is a good thing to require of evidence if evidence is to be a ground for knowledge, but when the aim is to assess more accessible probabilities to leverage to  $P(h)$ , high  $P(e)$  has no special advantage. And when a two-place relevance relation is at stake, we think a convincing argument has not yet been made.

**Acknowledgements** This study was supported by the UK AHRC 'Contingency and Dissent in Science' grant at LSE; we are grateful for this support. We also want to thank an anonymous referee as well as the participants in Jordi Cat's seminar at the University of Indiana for very helpful comments, and Sherri Roush for her inspiration.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

**Appendix**

**Theorem 1**  $LR > 1$  and  $P(h|e) > a > 0$  if and only if there exist  $x > 1$ ,  $1 > y > 0$  and  $0 < z \leq 1$  such that

- (i)  $LR \geq x$
- (ii)  $P(e) > y$
- (iii)  $P(e|h) \leq z$
- (iv)  $\frac{x - z/y}{x - 1} = a > 0$

*Proof* ‘if’

First, (i) and  $x > 1$  imply  $LR > 1$ .

Roush derives the following useful formula from the axioms of probability:

$$P(h|e) = \frac{LR - P(e|h)/P(e)}{LR - 1} \tag{1}$$

Solving for  $P(e)$  yields

$$P(e) = \frac{P(e|h)}{LR(1 - P(h|e)) + P(h|e)}$$

which with (ii) implies

$$\begin{aligned} \frac{P(e|h)}{LR(1 - P(h|e)) + P(h|e)} &> y \\ \Rightarrow P(e|h) &> yLR(1 - P(h|e)) + yP(h|e) \end{aligned}$$

which with (iii) implies

$$\begin{aligned} z &> yLR(1 - P(h|e)) + yP(h|e) \\ \Rightarrow [z/y - P(h|e)]/(1 - P(h|e)) &> LR \end{aligned}$$

Given (i), this then implies that

$$\begin{aligned} [z/y - P(h|e)]/(1 - P(h|e)) &> x \\ \Rightarrow z/y &> x(1 - P(h|e)) + P(h|e) \\ \Rightarrow z/y &> x - P(h|e)(x - 1) \\ \Rightarrow P(h|e)(x - 1) &> x - z/y \\ \Rightarrow P(h|e) &> (x - z/y)/(x - 1) \end{aligned}$$

Finally, (iv) then implies that

$$P(h|e) > a > 0$$

□

‘only if’

Let  $z = P(e|h)$  so (iii) holds. From (1) it follows that

$$P(h|e) = \frac{LR - z/P(e)}{LR - 1} \tag{2}$$

Define the following function

$$f(p,q) = \frac{p - z/q}{p - 1} \quad \text{for } LR \geq p > 1, \quad P(e) > q > 0.$$

Given the continuity of the right hand side of (2), as  $p \rightarrow LR$  and  $q \rightarrow P(e)$  then  $G(p, q) \rightarrow P(h|e)$ . Since  $P(h|e) > a$ , it follows by the definition of the limit there exist  $x$  and  $y^*$  such that  $LR \geq x > 1$  and  $P(e) > y^* > 0$  and

$$\begin{aligned} \frac{x - z/y^*}{x - 1} &> a \\ \Rightarrow x - z/y^* &> a(x - 1) \\ \Rightarrow x - a(x - 1) &> z/y^* \\ \Rightarrow y^* &> z/[x - a(x - 1)] \end{aligned} \tag{3}$$

Now define  $y$  by  $y = z/[x - a(x - 1)]$ .

Given  $x > 1, 0 < z \leq 1$  and  $0 < a < 1$  it follows that  $y > 0$ , and given (3)  $y^* > y$  follows. Since  $P(e) > y^*$  it follows that  $P(e) > y$ .

We have shown that  $LR \geq x > 1$  and  $P(e) > y > 0$  and  $y = z/[(1 - a)x + a]$ . Solving for  $a$  yields

$$\frac{x - z/y}{x - 1} = a$$

The result follows. □

**Corollary 1** *Given  $LR > 1, P(h|e) > a > 0$  if and only if  $P(e) > P(e|h)/[(1 - a)LR + a] > 0$ .*

*Proof* Let  $x = LR, z = P(e|h)$  and  $y = z/[(1 - a)x + a]$ . For any  $LR$  and  $P(e|h)$  (i) and (iii) are met and (iv) is met by definition of  $y$ . By the theorem therefore,  $P(e) > y$  if and only if  $P(h|e) > a > 0$ . The result follows from substitution of  $P(e|h)/[(1 - a)LR + a]$  for  $y$ . □

**Corollary 2** *Given  $LR > 1, P(h|e) > 1/2$  if and only if  $P(e) > 2P(e|h)/(LR + 1)$*

*Proof* Follows from Corollary 1 for  $a = 1/2$ . □

**Corollary 3**  *$LR > 1$  and  $P(e) > 1/[(1 - a)LR + a] > 0 \Rightarrow P(h|e) > a > 0$ .*

*Proof* Since

$$1 \geq P(e|h), P(e) > 1/[(1 - a)LR + a] > 0 \Rightarrow P(e) > P(e|h)/[(1 - a)LR + a] > 0,$$

the result then follows from Corollary 1. □

**Corollary 4**  $LR > 1$  and  $P(e) > 2/(LR + 1) \Rightarrow P(h|e) > 1/2$

*Proof* Follows from Corollary 3 for  $a = 1/2$ . □

**Theorem 2** Given  $LR > 1$ , for any  $x > 1$  there do not exist  $a$  and  $y$  such that  $P(h|e) > a$  and  $P(e) > y \Rightarrow LR > x$ .

*Proof* Solving (1) for LR yields

$$LR = [P(e|h)/P(e) - P(h|e)] / [1 - P(h|e)]$$

The right hand side is continuous in  $P(e)$  for any fixed value of  $P(e|h)$  and fixed non-unitary value of  $P(h|e)$ . Given this it follows that as  $P(e) \rightarrow P(e|h)$ ,  $LR \rightarrow 1$ . Therefore imposing restrictions  $P(h|e) > a$  and  $P(e) > y$  can not imply  $LR > x$  for any given  $x > 1$ , since one can always find a value of  $P(e)$  sufficiently close to  $P(e|h)$  such that  $x > LR > 1$  by the definition of the limit. □

**Theorem 3** Given  $LR > 1$ ,  $P(e|h)$  and  $P(e|\neg h)$  fixed,  $P(h|\neg e)$  strictly increases with  $P(e)$ .

*Proof*

$$\begin{aligned} P(h|\neg e) &= P(h \ \& \ \neg e) / P(\neg e) \\ &= [P(h) - P(h \ \& \ e)] / [1 - P(e)] \\ &= [P(h) - P(e)P(h|e)] / [1 - P(e)] \end{aligned}$$

But by Bayes theorem,  $P(h) = [P(h|e)P(e)]/P(e|h)$  so substituting

$$P(h|\neg e) = P(e)P(h|e) [1/P(e|h) - 1] / [1 - P(e)] \tag{4}$$

All the terms in the numerator increase, strictly increase or stay constant with increasing  $P(e)$  given fixed  $P(e|h)$ ,  $P(e|\neg h)$ , while the denominator strictly decreases. Therefore,  $P(h|\neg e)$  is a strictly increasing function of  $P(e)$ . □

**Theorem 4** Consider two possible posterior situations after updating using Jeffrey conditionalization on  $e$ . In one case one updates one’s degrees of belief on  $e$  to the posterior  $P_f(e)$ , in the other to  $P_f^*(e)$ , where  $P_f(e) < P_f^*(e)$ .

*Notation:*

- Let  $P_f(\bullet)$  denote the posteriors obtained by updating on  $P_f(e)$ .
  - Let  $P_i(\bullet)$  denote the priors before updating on  $P_f(e)$ .
  - Let  $P_f^*(\bullet)$  denote the posteriors obtained by updating on  $P_f^*(e)$ .
  - Let  $P_i^*(\bullet)$  denote the priors before updating to  $P_f^*(e)$ .
- If

- (a)  $P_f(e|h) = P_f^*(e|h)$
- (b)  $P_f(e|\neg h) = P_f^*(e|\neg h)$
- (c)  $LR = LR^* > 1$ .
- (d)  $P_i(e) = P_i^*(e)$ .

Then  $P_i(h) < P_i^*(h)$ .

*Proof* General result: in Jeffrey conditionalization  $P(h|e)$  and  $P(h|\neg e)$  remain unchanged on updating on  $e$ .

$$\begin{aligned} \text{So } P_i(h|e) &= P_f(h|e), \quad P_i(h|\neg e) = P_f(h|\neg e), \quad P_i^*(h|e) \\ &= P_f^*(h|e) \quad \text{and} \quad P_i^*(h|\neg e) = P_f^*(h|\neg e). \end{aligned}$$

By the axioms of probability:

$$P_i(h) = P_i(h|e)P_i(e) + P_i(h|\neg e)P_i(\neg e)$$

and

$$P_i^*(h) = P_i^*(h|e)P_i^*(e) + P_i^*(h|\neg e)P_i^*(\neg e)$$

Substituting it follows that

$$P_i(h) = P_f(h|e)P_i(e) + P_f(h|\neg e)P_i(\neg e)$$

and

$$P_i^*(h) = P_f^*(h|e)P_i^*(e) + P_f^*(h|\neg e)P_i^*(\neg e)$$

And by (d) it follows that

$$P_i^*(h) = P_f^*(h|e)P_i(e) + P_f^*(h|\neg e)P_i(\neg e)$$

Since  $P_f(e) < P_f^*(e)$ , by Roush's analysis it follows that  $P_f(h|e) < P_f^*(h|e)$  and by Theorem 3 that  $P_f(h|\neg e) < P_f^*(h|\neg e)$ . It follows by substitution of these inequalities into the above that  $P_i(h) < P_i^*(h)$ .  $\square$

## References

- Barnes, E. C. (2008). Evidence and leverage: Comment on Roush. *The British Journal for the Philosophy of Science*, 59, 549–557. doi:10.1093/bjps/axn022.
- Cartwright, N. (1979). Causal laws and effective strategies. *Nous (Detroit, Mich.)*, 13, 419–437. doi:10.2307/2215337.
- Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Clarendon Press.
- Cartwright, N. (2009, forthcoming). What is this thing called efficacy. In C. Mantzavinos (Ed.), *Philosophy of the social sciences. Philosophical theory and scientific practice*. Cambridge: Cambridge University Press.
- Howson, C., & Urbach, P. (2005). *Scientific reasoning: The Bayesian approach*. Chicago and La Salle, IL: Open Court Publishing.
- Roush, S. (2005). *Tracking truth: Knowledge, evidence and science*. Oxford: Clarendon Press.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.