Should evidence be probable? A comment on Roush

Damien Fennell & Nancy Cartwright

(Latest Draft: Jan 16th 2007, words including appendix: 8700)

0 Abstract

This paper argues that Sherrilyn Roush's (2005) dual claims that evidence should be highly probable and that this contradicts the Bayesian account of surprising evidence are flawed. It shows that evidence need not be highly probable, even by Roush's own criteria for evidence. It also provides a stronger argument for the demand that evidence have high probability from her own point of view by presenting an alternative characterization of evidence that also better meets her demand to connect evidence for a hypothesis with there being reason to believe the hypothesis.

1 Introduction

In her 2005 book Sherrilyn Roush offers a plausible set of conditions for when e is evidence for h, for dichotomous variables e and h. Roush has desiderata for evidence and she constructs her conditions for evidence to meet these. These include that evidence for h should discriminate between it and alternative hypotheses and that evidence for a hypothesis should be a good indicator of the hypothesis. These two conditions are also important for achieving another desideratum, that evidence should be a guide to belief in the hypothesis for which it is evidence, it should provide 'one who has it some reason to believe that the hypothesis is true' (p158).[1] She uses these desiderata to argue that 'ideally' evidence should be highly probable, which, she maintains, contradicts the Bayesian way to model 'surprising' evidence by low probability evidence. We shall argue both that her arguments in favour of high probability for evidence are weak and that her view does not contradict the Bayesian account of surprising evidence given her objectivist understanding of probability. We shall additionally provide a stronger argument for the demand that evidence have high probability from her own point of view by offering an alternative characterization of evidence to hers that, we claim, better meets her own demand connecting the existence of evidence with there being reason to believe.

2 How does Roush define evidence?

Roush argues 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) that LR > 1, where the likelihood ratio is defined by $LR = P(e|h)/P(e|\neg h)$. She also invokes a number of authors to argue that the likelihood ratio is also the best measure of how good evidence is at discriminating.

Roush takes the discrimination condition to be uncontroversial and focuses greater attention on the second condition for evidence, the *indication* condition. It requires

[1] Other desiderata are that the conditions for evidence should be given purely in terms of probability, that a measure of the support evidence provides a hypothesis should be given, that whether e is evidence should be an objective matter, and that the conditions for evidence should provide 'leverage' on the truth of the hypothesis, that is, knowing conditions for evidence are met should make it easier to know the truth of the hypothesis.

that the probability of the evidence and the likelihood ratio both be sufficiently high to ensure that P(h|e) > 0.5. This, of course, is equivalent to P(h|e) > 0.5.

The two conditions together define evidence. The definition of evidence is:

> e is *some evidence* for h if and only if
> ***FC1*** *(Discrimination Condition)*: LR > 1
> ***FC2*** *(Indication Condition)* : P(h|e) > 0.5

Roush also adds a stronger concept of evidence where the indication condition is more restrictive, it is:

> e is *good evidence* for h if and only if
> ***FC1***: LR > 1
> ***FC2'***: P(h|e) > b where b >> 0.5

We have labelled the conditions '***FC***' to point out that the expression of the conditions is ours and not Roush's. Roush offers a far more roundabout but exactly equivalent formulation:[2]

> *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 greater than 0.5'/'greater than some high threshold appropriate to having good reason to believe' (p.183).
>
> *R-addendum*: She adds '[w]hile on this view LR>1 is a necessary condition for … evidence, high P(e) is not necessary but is ideal' (p.183).

Roush's own formulation is odd for at least three reasons. First is the asymmetric treatment of ***FC1*** and ***FC2***. No constraints are placed on how ***FC1*** is to be satisfied. But ***FC2*** is to be met in a certain way. Second any values of P(e) and LR for which ***FC2*** hold are values 'sufficiently high' for it to hold, no matter how low they are. They may in fact both be extremely low yet P(h|e) be high. Even supposing ***FC1*** is met, P(e), we shall point out, can take any value consistent with high P(h|e). Perhaps this is why she adds that high P(e) is not necessary for evidence, presumably for evidence as understood by Roush, but only 'ideal'. But what then is the status of *R-addendum*? *R* is supposed to be a definition. Presumably by not adding the addendum into the definition Roush wishes to allow that e can be evidence, indeed *good* evidence, for h even if P(e) is low. Does she then have in mind three concepts: *some* evidence, *good* evidence, *ideal* evidence? It seems not since the addendum is not offered as a definition.

Roush explains that her roundabout formulation of ***FC2*** has two advantages. First she wants to 'leverage' the hypothesis, h, that is, she does not want criteria for evidence that would require, to be known, substantive knowledge of the hypothesis that the evidence is supposed to help obtain. Roush's more roundabout formulation of the indication condition, in terms of sufficiently high values of LR and P(e), has a leverage advantage when values for P(e) and LR can be inferred more easily than that

---

[2] Supposing P(e) > 0.

of P(h|e). The second reason for her more roundabout formulation of the condition is to highlight a disagreement she has with Bayesians and others about the size of P(e). The standard Bayesian view is that surprising evidence, represented by low P(e), raises degree of belief in a hypothesis by more than unsurprising evidence. In contrast Roush argues that evidence with high P(e) is more helpful for inferring the likely truth of a hypothesis.

3 What does Roush show?

Using the probability axioms Roush establishes

> ***A***. P(h|e) = [LR-P(e|h)/P(e)]/[LR-1].

She then 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/¬h) is also fixed.

Given these three conditions she points out that from ***A.***

> ***B.*** P(h|e) increases with P(e).

Roush elaborates on ***B.*** with a series of graphs of P(h|e) against LR for different possible values of P(e|h). Each graph assumes a different, fixed value of P(e). Roush uses the graphs to show that for high values of LR, a sufficiently high lower bound on P(e) implies P(h|e) greater than a half. Roush is particularly interested in lower bounds for P(e) and LR that are sufficient for the indication condition for some evidence (P(h|e) > 0.5). She shows that P(e) > 0.5 and LR > 3 are jointly sufficient for P(h|e) > 0.5. Although she does not herself state this result in this way, the general point that she uses the graphs to illustrate can be formulated as:

> ***C.*** For any b < 1, there exist d and c such that LR> d & P(e) > c => P(h|e) > b.

We prove this general result in the appendix.[3]

4 Roush's argument against surprising evidence

Roush argues that ***B***. and ***C.*** conflict with the Bayesian way of explaining why surprising evidence is better for raising the posterior probability of a hypothesis.[4] The Bayesian view can be illustrated from the standard version of Bayes Theorem.

> ***Bf.*** P(h|e) = P(h)P(e|h)/P(e)

In ***Bf.*** with P(e|h) fixed the posterior probability of h is raised more by evidence which has a low probability, i.e. the lower P(e) the greater the increase in P(h|e) from

[3] See Theorem 4 in the appendix.
[4] For an exposition of the Bayesian view see Howson and Urbach (2005, p.97) for example.

P(h). If surprising evidence is modelled by low P(e), this explains why surprising evidence has higher confirmation power for a hypothesis.

Roush argues against this widely accepted way of treating the higher confirmation power of surprising evidence. She argues instead – in line with ***B***. and ***C***. – that it is when evidence has a sufficiently high probability and LR is sufficiently high that the power of evidence is greater. In addition she nods towards an alternative to the Bayesian view of how to model surprising evidence in terms of P(e|¬h) instead of P(e), which could allow surprising evidence to provide higher confirmation of h while still allowing the 'ideal' condition that P(e) be high. Her independent argument for high P(e) is that if e is to be good evidence for h, e should provide good reason to believe h and that for e to be taken as reason for believing anything else, one must have reason to believe *it*. And for this P(e) should be sufficiently high because before one can accept e as evidence one must have reason to believe it. Otherwise how could it be the basis for revising any other beliefs?[5]

Both ***A***. and Bayes Theorem hold from the axioms of probability, so the conflict between Roush and the Bayesian views does not lie in the equations. The conflict is elsewhere. To identify its source it helps to look in more detail at Roush's argument.

5 A closer look at Roush's argument

Roush's graphical analysis shows that lower bounds on LR and P(e) are sufficient for a lower bound on P(h|e). Yet she concludes her graphical analysis with a 'proposal … that the second condition on evidence…be a lower bound on P(e)' (p.171). This makes explicit Roush's desire that high P(e) be a *necessary* criterion for evidence.

Despite Roush's proposal, a lower probability of evidence can make for better evidence on her own criteria. Consider for example the case where P(e|h) = 1,[6] which is one way to model 'h explains e' in the deductive-nomological account of explanation. In this case Bayes Theorem reduces to

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

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

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

and so

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

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

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

[5] In this way she rejects the Bayesian reading of P(e) as giving the probability of e before it is observed.
[6] Similar examples can be generated for any fixed non-zero value of P(e|h).

In this case lowering P(e) by lowering P(e|¬h) simultaneously produces improvements both in LR and in P(h|e). Lowering e in this way makes e better evidence for h on both Roush's criteria. This is despite Roush's graphical analysis, which shows a sufficient but not a necessary condition. It is true, as she concludes from her graphs, that 'increasing P(e) with fixed or rising LR will have the effect of increasing P(h|e)' (p.168). But it is equally true that decreasing P(e) with rising LR can have the effect of increasing P(h|e). So 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 ***FC1*** and ***FC2*** can be maximally satisfied while P(e) takes on any value whatsoever. For suppose e is a perfect sign of h; i.e. e ≡ 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.

Besides saving the venerable connection between evidence and surprise this example has another nice result as well. Whenever there are two independent criteria for the same thing questions of trade-off come up. What follows when one criterion improves at the cost of the other? But here trade-off is avoided. In this case (or any case with fixed P(e|h)) when the results implied by a hypothesis become more surprising, the results can become better evidence by both criteria at once. Thus, it seems the more surprising evidence is (in terms of lower P(e)), the better evidence e can be in terms of Roush's criteria for evidence.

The example above shows how e for which P(e) is low can be better evidence according to Roush's definition. It appears to lend support to the conventional view that surprising evidence has higher confirmation power than unsurprising evidence where surprising evidence is modelled by low P(e). It also brings out the source of the disagreement between Roush's and the Bayesian view of surprising evidence. The difference is that Roush considers the impact on P(h|e) of increases in P(e) for a *fixed* LR. In contrast the Bayesians – and the example presented here – do not assume a fixed LR. By allowing LR to increase with lower P(e), lower P(e) can raise P(h|e) in line with the Bayesian view of surprising evidence.

We should also note that Roush's graphical arguments for high P(e) depend heavily on the asymmetry with which she treats the two independent criteria for evidence, an asymmetry that we remarked on in section 2. Suppose e is 'candidate' evidence for h in the sense that ***FC1*** is well satisfied (i.e. LR is high). Then it is true that high P(e) is sufficient for the satisfaction of ***FC2*** (i.e. high P(h|e)). But the exactly symmetric claim is not true. Suppose e is 'candidate' evidence for h in the sense that ***FC2*** is well satisfied (i.e. P(h|e) is high). Then it is not true that high P(e) is sufficient for the satisfaction of ***FC1*** (i.e. LR is high). We prove this in the appendix.[7] The result illustrates an important asymmetry in Roush's analysis that, although a high P(e) is useful for obtaining the high P(h|e) when LR is sufficiently high, a high P(e) is not sufficient for high LR when P(h|e) is high. Yet there is no special reason for considering either criterion differently from the other. So again it seems that formula ***A.*** and the associated graphical analysis do not provide a very good argument for taking high P(e) as the ideal.

[7] See Theorem 5 in the appendix.

Roush defends her claim that P(e) be high in order for e to be evidence for anything on three fronts. The first involves the arguments based on formula ***A.*** and the accompanying graphs. This, we have just argued, provides very 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 central arguments on this front is that Bayesians themselves need high P(e) to warrant the method they recommend for belief revision. We think this argument rests on a simple mistake, which we will discuss in section 8 when we talk about Bayesian conditionalization. She also hopes to model the surprisingness of evidence by $P(e|\neg h)$ – a proposal she describes but proposes to develop elsewhere[8] – so that counting surprising evidence as good evidence does not count against demanding P(e) be high. In addition on this front she 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 1. Finally she provides 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 are arguments that show either that P(e) need not be small or that it is no harm for it to be big. These are exactly in line with the view that follows from her own definition of evidence, that the probability of e is irrelevant to whether e is evidence or not. A positive argument is still required for the claim expressed in the odd add-on to the definition, that 'high P(e) … is ideal'. That seems to be left to her third line of defense.

The positive argument for high P(e) is woven in and out among the discussion of the Bayesian account of surprising evidence and elsewhere and has little explicit statement. But here are a few: 'it would, I think, be inappropriate to say that a subject "has" evidence e when she does not have a belief that e' (p.153). Or '… in a case where I have a very low degree of belief that e occurred, what reason could there be to think that e is evidence of anything for me?' (p.173). Or later '…if the indication condition is fulfilled in the ideal way with P(e) high, that will tell us that e is evidence for *something*' (p.176). The idea seems to be that nothing counts as evidence unless there is more reason to believe it rather than its opposite. But why?

One ground for this seems to be her own central idea that having evidence gives one reason to believe: if I didn't believe e, how could e give me reason to believe anything else? This looks, however, more like a verbal slip than an argument since the task Roush undertakes is to define 'e is evidence for h', not what it is for a subject to have evidence for h. We could easily agree that when I *have* evidence e for h then I *have* reason to believe h. So in order for me to *have* e, my assessment of P(e) should be sufficiently high; I can't use e as a reason for h unless P(e) is high for me. This can be perfectly true without jeopardizing the antecedent idea that ***FC1*** and ***FC2*** characterize a notion of what makes e evidence for h. Indeed this fits nicely with Roush's own idea of leverage. If I want to assess the truth of h and I am fairly ignorant of the facts surrounding it, I don't want to start collecting information at random since collecting information is costly. I would like some criteria – a characterization of evidence – that tells me what facts would be reasons for h. Then I go out and try to ascertain the probability of those. In this discussion we have supposed that it is this antecedent notion of *evidence* that Roush is trying to

[8] See Roush (forthcoming).

characterize. One can then of course add as a canon of good reason: do not believe h on the basis of e alone unless P(e) is high.

We conclude then that Roush's ground for high P(e) as an ideal for something to count as evidence is not strong on any of the three fronts we have identified. There is, however, another strategy that starts with Roush's own insistence that evidence provides reason to believe and makes a stronger case for high P(e). Indeed when P(e|h) = 1 (which we noted is one standard way to model deductive-nomological explanation) P(e) must be greater than ½. The strategy involves strengthening FC2, which we shall argue should be done anyway to express Roush's demands better. We turn to this strategy now.

6 Evidence and reason to believe the hypothesis

A key desideratum for Roush is that evidence should provide reason to believe a hypothesis when the evidence is taken into account. This desideratum, 'reason to believe', is supposed to be met in Roush's definitions of evidence by the indication condition. For some evidence P(h|e) > 0.5 must hold. As Roush sees it, this ensures that one has some reason to believe h (rather than its alternative ¬h) when one has evidence e. In the case of good evidence it must be the case that P(h|e) > b for some large b appropriate to one's threshold for having reason to believe h. As she puts it, '… 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)] is greater than 0.5' (p.165).

In this section we argue that the desideratum, P(h|e) > 0.5, is not sufficient for there to be reason to believe in the hypothesis when the evidence is less than certain. This has implications for Roush's theory of evidence. Addressing the problem requires either an elaboration on what it means to *have* evidence, or a strengthening of Roush's second criteria for evidence to P(h) > 0.5.

In her paper Roush presents the example of Mary (p.171-172) to illustrate how high P(e) and high LR are sufficient for high P(h|e). Here we consider the similar case of John, who may also be suffering from a disease, D. John is suffering from a number of symptoms that suggest a blood test for a marker, d, related to the disease is appropriate. We let

h: John has disease D.

e: The marker d is in John's blood.

Suppose it is known that in a population of people like John when a patient has disease D then d is also present, so P(e|h) = 1 is known. Suppose also that condition d is uncommon when the disease is absent, specifically suppose that it is known that LR = 3. Now suppose John is given the test 20 times and it comes up positive 13 times, from which it is inferred that P(e) = 0.65. Given this one can calculate that P(h|e) = 0.73 using ***A***.. This example exactly parallels Roush's Mary example though differing in the numbers just enough to make our point about P(h) versus P(h|e). It is a case where e is some evidence for h and if one's threshold for good evidence is 0.73 or less then e is good evidence for h.

Following the way Roush discusses the Mary example, since e is some evidence for h, e provides some (and possibly good) reason to believe h in the case of John. But just what does this mean in the John example? For Roush this is formalised by P(h|e) being sufficiently high. And it is clear that this condition is met in both the Mary and John cases. But in the case of John, if all we know is exactly what has been stated, do we have reason to believe h? It seems difficult to argue this since it is easily calculated that P(h) = 0.475.[9] Therefore, in John's case the correct inference is that he is more likely *not* to have the disease. So it seems counterintuitive to claim that e gives us reason to believe John has the disease D because we have more reason to believe that John does not have the disease.

More generally whenever LR > 1 (the discrimination condition for evidence) it follows that P(h|e) > P(h). Thus, a lower bound on P(h|e) need not imply a sufficiently high P(h), even to a level where the hypothesis can be inferred to be more likely than its alternative.

Of course this example appeals to an intuition that a lower bound on P(h) is a more natural criterion for there being reason to believe in h than a lower bound on P(h|e), as adopted by Roush. This intuition can be justified further by looking carefully at what P(h|e) must represent in Roush's framework. Unlike the Bayesians Roush explicitly adopts objective probabilities. Therefore, despite sometimes slipping into Bayesian language of priors and posteriors at points in the chapter, P(h) and P(h|e) cannot be prior and posterior probabilities of the hypothesis, h. P(h) is simply the probability the hypothesis is true for an arbitrary member of the population of interest and P(h|e) is the probability the hypothesis is true for an arbitrary member of the sub-population in which e is true. In the John and Mary examples P(h) is the probability that an individual with symptoms like those exhibited by John (or Mary) has the disease. In contrast P(h|e) is the probability that an individual with symptoms like those of John (or Mary) who also has the marker in their blood has the disease. But, as Roush herself emphasises,[10] P(e), the probability that John (or Mary) has the marker, need not be one. When P(e) is not one, P(h|e) and P(h) diverge. Moreover, in the cases Roush is interested in, P(e) is generally not one. Given that P(h|e) is the probability of the hypothesis if e is true, it seems counterintuitive to adopt this as giving reason to believe h when the evidence is less than certain.

There is an immediate reply that may seem natural. If in cases like the John and Mary examples, a sufficiently high threshold on P(h) is what is desirable for having reason to believe h, then what role does evidence play? It may seem that we are ignoring the useful information evidence provides, since we are focusing on the probability of h in the whole population independently of whether or not evidence is true for the different members of the population. There is a sense in which this is correct. In the John case P(h) gives the probability that an individual like John (i.e. with the same symptoms, with the same dispositions in relation to the blood test) has the disease. But evidence still plays a role, the reason to give John the blood test is to infer P(e) for this population. Yet it is an inference *for John* only in the sense that it is assumed that he is randomly drawn from this population. The inference of P(e) is then useful because it allows one to infer P(h) for the population of John-like individuals via Roush's formula ***A***.. This is then useful for making inferences about whether John has the disease.

---

[9] Since P(e|h) = 1, P(h) = P(h|e)P(e) = (0.73)(0.65) = 0.475.

[10] Or as Roush puts it in her Mary example: 'one positive test does not a positive tester make: it could have been a fluke' (p.171).

So in the John case, e is evidence – following Roush's two criteria – that John has the disease, but e fails to provide reason to believe that John has the disease. Therefore, it fails to meet Roush's desideratum for reason to believe.

Now one might retort that in the John case that e nevertheless provides reason to believe the hypothesis if one *had* it. This re-interprets the John example as a case where there is a failure to have evidence rather than one where there is a failure of the evidence to provide reason to believe the hypothesis. This also reads Roush's comments that the evidence should provide reason to believe the hypothesis *when it is taken to account*[11] as a requirement that the evidence would give reason to believe the hypothesis were it true.[12] However, the problem remains that the evidence in this case implies that John is more likely not to have the disease. If one attributes this problem to a failure to have the evidence, and one wants it to be such that in a case where the evidence were had – so to speak – the hypothesis would be more likely than not (unlike the John example), then one needs an argument that shows that evidence provides reason to believe the hypothesis when e is evidence and when one has the evidence. Roush does not provide such an argument. Perhaps a proof can be given that shows that the hypothesis is probable when e is evidence and it is sufficiently probable (taking this as what it means to have the evidence). However, in the appendix we prove that if e is some evidence then $P(e) > ½(P(e|\neg h) + P(e|h)) \Leftrightarrow P(h) > ½$.[13] This is worrying since the lower bound on the probability of evidence, sufficient for the hypothesis to be probable, depends on the hypothesis. This is unsatisfactory if one assumes, intuitively, that 'having' a proposition should depend solely on how likely that proposition is. So having the evidence should not depend on its relationship with the hypothesis, as it does in this simple attempt to provide a proof to support this response to the problem in the John example. Therefore, further work seems to be required to make this approach work.

An alternative approach to deal with the problem is to strengthen the definition of evidence. This approach takes it that evidence e gives one reason to believe h, that is, ensures $P(h) > 0.5$, in virtue of it being evidence. This move, unlike the approach above, has the additional advantage that it implies that evidence must be sufficiently probable for it to be evidence. We now consider this option.

7 A possible revision to Roush's definition of evidence?

Given the discussion above a natural revision to Roush's desideratum for having reason to believe is.

> There is some reason to believe h if and only if $P(h) > 0.5$
> &
> There is good reason to believe h if and only if $P(h) > b$, where $b>>0.5$.

---

[11] Roush uses this language when describing the desideratum of having reason to believe. She writes 'I will also identify a second kind of condition that … puts a lower bound on the posterior probability…, that is, the probability that the hypothesis is true once the evidence is *taken into account*.' (p.158, emphasis added).

[12] This makes sense of the criterion that $P(h|e) > 0.5$.

[13] See theorem 2 in the appendix.

If this revised desideratum is to be met by Roush's theory of evidence, then the indication condition needs to be strengthened to $P(h) > 0.5$ and $P(h) > b$ respectively for some and good evidence. That is, no e can be evidence for h unless there is some reason to believe h, as characterized above.

But this is obviously problematic, since it undermines the interpretation of the second condition for evidence as one of the evidence being a good indicator. Consider $P(h|e) > 0.5$ which means that when the evidence is true, then h is likely to be the case. This is a plausible formalisation of what it is for e to indicate h. In contrast $P(h) > 0.5$ simply means the hypothesis is more likely to be true than not in the population. It is completely independent of e, and thus cannot bear an interpretation that it is an indication condition *for e*. This suggests a tension between two of Roush's desiderata for evidence, between her desire to capture the indicativeness of evidence in her criteria for evidence and her desire that these imply that if there is evidence for h then there is reason to believe h.

Despite this tension the way out may simply be to adopt the stronger $P(h) > 0.5$ as the second condition for evidence and to drop its indication interpretation. Though this is less satisfactory for the indication desideratum, not all is lost, since when the discrimination condition is met ($LR>1$) then $P(h|e) > P(h)$. Thus $P(h) > 0.5$ implies the old indication condition $P(h|e) > 0.5$. So, although the second condition has lost its indication interpretation, the indication property of evidence still follows from a stronger definition of evidence.

Moreover one can show that there are lower bounds for LR and P(e) which are sufficient for a lower bound on P(h) and develop a counterpart 'roundabout' formulation as Roush did for P(h|e).[14] All of this suggests the following revised, stronger definition of some evidence.

> e is some evidence for h if and only if
> ***FC1***: $LR > 1$
> ***FC2**** : $P(h) > 0.5$.[15]

A similarly stronger definition of good evidence could be constructed. Also, since $LR>1 \Rightarrow P(h|e) > P(h)$, these revised definitions of evidence imply Roush's original counterpart definitions.

This revised definition has another advantage over the original definition of evidence. It provides better support for Roush's argument on surprising evidence. To see this advantage, reconsider the earlier example, where increasing LR and decreasing P(e) by lowering $P(e|\neg h)$ led to an increase in P(h|e). The case was problematic for Roush because it showed that one need not have high P(e) for e to be some evidence. It was

---

[14] We prove this in the appendix as Theorem 3.

[15] There may be a mild concern that the revised requirement for e to evidence for h lays down a demand on h alone, not just on some relations between e and h. This should not be a problem from Roush's point of view, though, since she wants to do the same with e. Recall that fulfilling the indication condition with P(e) high 'tells us that e is evidence for *something*' (p.174, italics original). Similarly here one can say that fulfilling ***FC2**** shows that h can evidenced by something. In any event, as proved in the appendix, ***FC2**** can be replaced by $P(e) > \frac{1}{2} P(e|h) + \frac{1}{2} P(e|\neg h)$, since when ***FC1*** holds, this is equivalent to $P(h) > \frac{1}{2}$.

also a case where better evidence was achieved on *both* criteria by *lowering* P(e) by lowering P(e|¬h). The revised definition has the advantage that lowering P(e) in this way need not raise P(h).

To see this reconsider the example presented earlier. There it was assumed that P(e|h) = 1 and it was derived that

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

and

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

Thus, lowering P(e) by lowering P(e|h) leads to an increase in P(h|e). But what about P(h)? Well in deriving the increase in P(h|e) it was assumed that P(h) was fixed. So, although lowering P(e) by lowering P(e|¬h) led to an increase in P(h|e) in the example, it was not accompanied by a change in P(h). This shows that requiring the stronger second condition on evidence, that is, a lower bound on P(h) rather than on P(h|e), weakens the criticism presented above of Roush's theory of evidence, since with the revised definition of evidence, an increase in P(e) by lowering P(e|h) need not increase P(h).

More generally, with the revised definition of evidence, it is possible to have a lower P(e) and a lower P(e|¬h) with either a decrease, increase or no change in P(h). This can be seen – again restricting ourselves for simplicity to the case where P(e|h) = 1 – from the following, easily derived equation for P(h).

H. $P(h) = [P(e) – P(e|\neg h)]/[1-P(e|\neg h)]$

Here lowering P(e) and P(e|¬h) may lower P(h), where the decrease in P(e) is sufficiently greater than that in P(e|¬h). A decrease in P(e), provided it is offset by a sufficiently large decrease in P(e|¬h), can also lead to an increase in P(h).[16] Also ***H.*** implies that $P(h) \leq P(e)$. Therefore, a lower P(e) implies a lower maximum possible value of P(h).

Importantly $P(h) \leq P(e)$ implies $P(e) > ½$ given ***FC2**** . So in the case where P(e|h) =1, $P(e) > ½$ is necessary for e to be evidence under the stronger, revised definition of evidence. A similar result holds generally. Even when P(e|h) is not one, ***FC2**** implies a non-trivial lower bound on P(e). In the appendix we prove that for LR > 1, $P(h) > a \Leftrightarrow P(e) > a P(e|h)+(1-a)P(e|\neg h)$.[17] So for the revised definition of evidence, since $P(h) > ½$ it is *necessary* that $P(e) > ½P(e|h) + ½P(e|\neg h)$. Thus, unlike Roush's definition of evidence, the stronger definition has the desirable feature – from Roush's perspective – that evidence must be sufficiently probable in order for it to be evidence.

Nevertheless, these results do not imply that lower P(e) is never better for achieving high P(h). As noted above, it is possible to lower P(e), lower P(e|¬h) and increase P(h). So, as with Roush's definition of evidence, it is possible to raise P(h) and LR by lowering P(e) and lower P(e|¬h). However, unlike Roush's definition, in the stronger

[16] For example when P(e) = 0.955 and P(e|¬h) = 0.9, then P(h) = 0.55. If P(e) is lowered to 0.9 and P(e|¬h) lowered to 0.75, then P(h) increases to 0.6.
[17] See theorem 2 in the appendix.

definition, lowering P(e) and P(e|¬h) need not lead to an increase in P(h). In addition, since P(e) is bounded from below by ½P(e|h)+½P(e|¬h), one cannot lower P(e) 'too much'. Thus, the revised second condition for P(h) implies that evidence must be sufficiently. It also avoids the problematic case for Roush's theory of evidence in which lowering P(e) by lowering P(e|¬h) necessarily leads to better evidence. In sum the revised definition of evidence appears preferable from a Roushian perspective.

8 Keeping Roush's and the Bayesian approach separate

The revised definition of evidence combined with the proposed new condition for having reason to believe also helps to clarify an important difference between Roush's approach to evidence and the Bayesian approach to confirmation.

The revised definition of evidence takes as a key desideratum that evidence imply a sufficiently high probability of the hypothesis. Therefore, in the situations in which there is evidence the probability of the hypothesis will be high. This will also tend to be the case for Roush's theory of evidence (in spite of the John example above) because it can be shown that a lower bound on P(h|e) also implies a lower bound on P(h).[18] In contrast in the Bayesian approach P(h) need not be high, it represents the prior probability of the hypothesis, the degree of belief in a hypothesis given some limited prior information. This may well be low and indeed it will be wherever hypotheses are initially taken to be unlikely.

To analyse the differences with Bayesianism, consider the following simple schema for how Bayesians interpret belief revision. Let P denote the prior probability, P' the posterior probability after observing e. Initially the observers' degrees of belief are given by P(h), P(h|e), P(e)…. Now suppose the observers observe e. At this point having observed e the observers believe e is the case and update their degree of belief in e to 1, so that P'(e) = 1. How do the observers update their other beliefs? Well the observers adopt as posterior probabilities, their prior probabilities on all other statements conditioned on e. Thus, updating for any proposition f follows the rule:

$$P'(f) = P(f|e)$$

This is reasonable because these prior probabilities represented the observers' prior degrees of belief if e was the case. Now that e is seen to be the case, observers replace their priors with posteriors equal to their priors conditioned on e. In this way Bayes Rule for updating given by

$$P(f|e) = \frac{P(e|f)P(f)}{P(e)}$$

is seen naturally to make reference *only* to prior probabilities. It is simply mistaken to read P(e) in the formula as needing to refer to the posterior probability of e, as Roush argues by claiming that P(e) must be high even in the Bayesian analysis. Simply put this is because the act of adopting a high degree of belief in e on observing e is itself an updating act, where P'(e) is set to P(e|e). Unfortunately it appears that Roush is mixing her view, where P(e) ideally should to be high for it to be evidence for anything, with the Bayesian view where P(e) is the degree of belief in e prior to observing it.

[18] See theorem 1 in the appendix.

This can be seen in Roush's contrasting description of the process of Bayesian conditionalization. She says 'your degree of belief in e prior to the conditionalization is just P(e), so high P(e) is (almost) sufficient for you to take e as evidence for whatever happens to be positively relevant, that is, to conditionalize upon it … If … P(e) is your degree of belief as far back as before you observe e, then you have no justification for strict conditionalization on e because you do not have confidence that e is true. 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' (p. 174).

It is as if Roush supposes that Bayesians have a three-step process. Observers begin with degrees of belief represented by the 'antecedent' probability P. 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 in the third step, changing their degrees of belief to those represented by the 'posterior' probability P', where for any f, P'(f) = P(f/e). But of course Bayesians do not take three steps but only two. They observe e at the first step and at the second, revise their probabilities in one fell swoop to P', which among other features sets the probability of e to 1. Indeed for a Bayesian, nothing else is possible: Bayesians – as Roush herself characterizes them – always assign degrees of belief to every well-formed formula so one cannot just revise one's degree of belief in one fact and *then* conditionalize to set the rest of one's degrees of belief.

This is the usual and familiar story, a story that puts justification at a different point from where Roush proposes to put it. For the Bayesian the new probability is justified by the observation of 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.[19]

The revised definition of evidence also keeps clear of confusion with the Bayesian reading by replacing the desideratum of having reason to believe the hypothesis with high P(h). Roush's desideratum of a high P(h|e) not only invites confusion with the Bayesian desideratum of a high posterior probability of h but also seems misguided given her interpretation of probabilities. This was shown by the John example above.

Importantly the revised definition of evidence and the Bayesian approach use the same probability calculus but do *not* directly conflict.[20] In the revised definition of evidence the desirable case is that in which P(e) and LR are sufficiently high to place a lower bound on P(h). Though high P(e) is not necessary for this, it certainly helps, for essentially the reasons that Roush presents. By contrast in the Bayesian approach the aim is to update on e's to confirm/disconfirm the hypothesis. In this case, as the Bayesian's examples of surprising evidence and the counter-example presented above (for low P(e) to be necessary for some evidence) show, low P(e) can undoubtedly be helpful. In this case the low P(e) tends to be associated with a low P(h) and a high

[19] Roush also mentions Jeffrey conditionalisation. But she doesn't discuss it so neither will we.
[20] They obviously conflict in many other ways, in their interpretations of inferences and probabilities for instance. The point here is that the conflict does not lie in the formalisations, as Roush appears to suggest by arguing from formal considerations to claims that P(e) ought to be high for high P(h|e).

P(h|e), which fits precisely with Bayes rule. But here, unlike in the Roush or revised framework, low P(h) is not an issue. It reflects initial scepticism about h, which is revised in the face of an unexpected e, represented by low P(e). Just as one would expect when one is surprised by the evidence.

9 Conclusion

Roush's arguments for high P(e) if e is to be 'ideal' evidence that are based on formula ***A.*** and the associated graphs are weak. They rely on treating the two criteria she offers for evidence asymmetrically for no good reason and they show only that high P(e) is sufficient for the second criterion to be met given that the first is, not that it is necessary. Both criteria can be met maximally and P(e) still take any value at all. On the other hand, her independent motivation for high P(e) is better met, we have argued, by insisting that for any e to be evidence not only must P(h|e) be high, so too must P(h). This fits far better with her demand that evidence provide reason to believe.

In addition it seems that both Roush's approach suitably revised and the Bayesian approach make sense of inference in the face of evidence. The apparent tension between the two approaches, exploited by Roush to criticise the Bayesian analysis, is illusory. The differences between the two approaches are not at the level of the probability calculus but lie in the distinctive interpretations of probabilities and inference the two approaches adopt. As a result Roush's criticisms of the Bayesian analysis are flawed. They appear to rest on her adoption of a high P(h|e) as reason to believe and a high P(e) as sufficient (but not necessary) to ensure this given high LR. As made clear in this paper, high P(h|e) is more properly the desideratum for Bayesian analysis, but in this case, low P(e), as shown by the example in section 5, can help with this goal. In contrast, once one moves to high P(h) as the desideratum for the Roushian treatment of evidence, high P(e) is more plausibly helpful for evidence to give reason to believe.

Appendix

Before proving the relevant results, first note two useful formulae.

In her chapter Roush derives the following useful formula from the axioms of probability:

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

By Bayes Theorem P(e|h)/P(e) = P(h|e)/P(h) substituting this into (1) yields:

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

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

So

$$P(h|e) = \frac{LR}{[LR - 1 + 1/P(h)]}. \qquad \ldots \qquad (2)$$

We can then prove:

Theorem 1: If LR > 1 then for any 0 < b < 1 the following are equivalent:

(i) P(h|e) > b
(ii) P(h) > b/[b + LR(1 - b)]
(iii) P(e) > P(e|h)/[b + LR(1 - b)]

Proof:

Substituting into (1) P(h|e) > b is equivalent to

$$\frac{LR - P(e|h)/P(e)}{LR - 1} > b$$

Since LR >1, this is equivalent to

$$LR - P(e|h)/P(e) > b(LR - 1)$$
$$\Leftrightarrow \quad P(e|h)/P(e) < LR - b(LR - 1)$$
$$\Leftrightarrow \quad 1/P(e) < [LR - b(LR - 1)]/P(e|h)$$
$$\Leftrightarrow \quad P(e) > P(e|h)/[LR - b\,(LR - 1)]$$
(since 0<b<1, LR > 1 => LR- b(LR - 1) > 0)
$$\Leftrightarrow \quad P(e) > P(e|h)/[LR(1-b) + b]$$

So (i) <=> (iii)

Now by (2), P(h|e) > b is equivalent to

$$P(h|e) = \frac{LR}{[LR - 1 + 1/P(h)]} > b$$

Given LR > 1, this is equivalent to

$LR > b[LR - 1 + 1/P(h)]$

$\Leftrightarrow$ $LR - b(LR - 1) > b/P(h)$

$\Leftrightarrow$ $P(h)/b > 1/[LR - b\,(LR - 1)]$

$\Leftrightarrow$ $P(h) > b/[LR(1-b) + b]$

So (ii) <=> (iii) ■

Corollary (Reparametrizations of Theorem 1): If LR > 1, then for any 0 < a < 1 (iv), (v) and (vi) are equivalent and for any 0 < c < 1 (vii), (viii) and (ix) are equivalent where

(iv) $P(h|e) > a/[a + (1 - a)/LR]$
(v) $P(h) > a$
(vi) $P(e) > P(e|h)[a + (1 - a)/LR]$

(vii) $P(h|e) > [LR - P(e|h)/c]/[LR - 1]$
(viii) $P(h) > [cLR/P(e|h) - 1]/[LR - 1]$
(ix) $P(e) > c$

Proof: To prove (iv) <=> (v) <=> (vi), let $a = b/[b + LR(1 - b)]$, solve for b in terms of a, use this to substitute a for b in Theorem 1, result follows. Likewise to prove (vii) <=> (viii) <=> (ix), let $c = P(e|h)/[b + LR(1 - b)]$, solve for b in terms of c and use this to substitute c for b in Theorem 1. The result follows. ■

Theorem 2: If LR > 1 then $P(h) > a \Leftrightarrow P(e) > a\,P(e|h) + (1-a)P(e|\neg h)$.

Proof: From the corollary if LR > 1 then $P(h) > a \Leftrightarrow P(e) > P(e|h)[a + (1 - a)/LR]$, since $LR = P(e|h)/P(e|\neg h)$, so $P(h) > a \Leftrightarrow P(e) > a\,P(e|h) + (1-a)P(e|\neg h)$. ■

Theorem 3: If LR > d > 1 and P(e) > c for c such that 0 < c < 1, then

(a) $P(h|e) > [cd - 1]/[c(d - 1)]$
(b) $P(h) > [cd - 1]/[d - 1]$

Proof: By the corollary $P(e) > c \Rightarrow P(h|e) > [LR - P(e|h)/c]/[LR - 1]$

The right hand side is equivalent to

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

$\Rightarrow$ $P(h|e) > 1- [1/c - 1]/[LR - 1]$ (since $P(e|h) \leq 1$)[21]

$\Rightarrow$ $P(h|e) > 1 - [1/c - 1]/[d - 1]$ (since LR > d)[22]

[21] Because $P(e|h) \leq 1 \Rightarrow P(e|h)/c - 1 \leq 1/c - 1 \Rightarrow [P(e|h)/c - 1]/[LR - 1]c \leq [1/c - 1]/[LR - 1] \Rightarrow 1 - [P(e|h)/c - 1]/[LR - 1]c \geq 1 - [1/c - 1]/[LR - 1]$.

<=>   P(h|e) > [cd - 1]/[cd - c]

So LR > d and P(e) > c => P(h|e) > [cd - 1]/[c(d - 1)]

This proves (a), the proof for (b) is similar:

By the corollary P(e) > c => P(h) > [cLR/P(e|h) - 1]/[LR - 1]

=>	P(h) > [cLR - 1]/[LR - 1]	(since P(e|h) ≤ 1)[23]
<=>	P(h) > [c - 1/LR]/[1 - 1/LR]	
<=>	P(h) > 1 - [1 - c]/[1 - 1/LR]	
=>	P(h) > 1 - [1 - c]/[1 - 1/d]	(since LR > d)[24]
<=>	P(h) > [c - 1/d]/[1 - 1/d]	
<=>	P(h) > [cd - 1]/[d - 1]	

So LR > d and P(e) > c => P(h) > [cd - 1]/[d - 1] which proves (b).

Theorem 4: For any 0 < b < 1, there exist 0 < c < 1 and d > 1 such that LR> d & P(e) > c => P(h|e) > b.

Proof: For any 0 < b < 1, d > 1, define c by [cd - 1]/[d - 1] = b, solving for c yields c = 1/d(1-b) + b. Since, 1/d < 1, and 0<b<1, it follows that 0 < c < 1. Substituting b for [cd - 1]/[d - 1] in Theorem 3, it follows that for any 0 < b < 1, there exist 0 < c < 1 and d > 1 such that LR> d & P(e) > c => P(h|e) > b . ■

Comment: This result and result (a) of theorem 3 are what Roush illustrates with graphs in the chapter, that a lower bound on LR (greater than or equal to 1) and a lower bound on P(e) are sufficient for a lower bound on P(h|e). Result (b) of theorem 3 extends this, showing that the lower bounds on LR and P(e) are also sufficient for a lower bound on P(h).

Theorem 5: If LR > 1, then for any d > 1, then there do not exist b and c such that P(h|e) > b and P(e) > c => LR > d.

Proof: Solving (1) for LR yields

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

And from this it is easily checked that for any fixed non-zero value of P(h|e) as P(e) tends to P(e|h) LR tends to 1. Therefore imposing restrictions P(h|e) > b and P(e) > c can not imply LR > d for d > 1, since one can always find a value of P(e) sufficiently close to P(e|h) so that 1 < LR ≤ d (follows from definition of the limit). ■

---

[22] Because LR > d => LR -1 > d - 1 => [1/c - 1]/[LR - 1] < [1/c - 1]/[d - 1] => 1 - [1/c - 1]/[LR - 1] > 1 - [1/c - 1]/[d - 1].

[23] Because P(e|h) ≤ 1 => 1/P(e|h) ≥ 1 => cLR/P(e|h) ≥ cLR => cLR/P(e|h) - 1 ≥ cLR - 1 => [cLR/P(e|h) -1]/[LR - 1] ≥ [cLR - 1]/[LR - 1].

[24] Because LR > d => 1/LR < 1/d => 1 - 1/LR > 1 - 1/d => [1 - c]/[1 - 1/LR] < [1 - c]/[1 - 1/d] => 1 - [1 - c]/[1 - LR] > 1 - [1 - c]/[1 - 1/d].

References

Howson, C. and Urbach, P. (2005) *Scientific Reasoning: The Bayesian Approach*, 3rd edition, Chicago and La Salle, Ill.: Open Court Publishing.

Roush, S. (forthcoming) *The Probability of Evidence*.

Roush, S. (2005) *Tracking Truth: Knowledge, Evidence and Science*, Oxford: Clarendon Press.