

“Evidence and Leverage: Comment on Roush”
 Eric Christian Barnes
 2006 Pacific APA: Author Meets Critics Session

The central thesis of Sherri Roush’s excellent book *Tracking Truth: Knowledge, Evidence and Science* (2005 Oxford University Press) is that a person has knowledge if her beliefs track the truth. This is to say, roughly, that if a statement is true she is likely to believe it, and if it is false she is unlikely to believe it. But as Roush explains this plausible picture of knowledge carries implications for the nature of evidence – intuitively, e is evidence for h if e tracks h : that is, e is likely to be true if h is true, and e is likely to be false if h is false. On this approach, a natural procedure for estimating the degree to which e supports h is in terms of what is known as the likelihood ratio (LR): $p(e/h)/p(e/\sim h)$. Intuitively the LR measures the degree to which e tracks h : when this ratio is greater than 1, then e confirms h – the higher the ratio goes, the more strongly e tracks (and thus confirms) h . This is because the LR goes higher the more probable e is assuming h is true, and the less probable e is assuming h is false. LR offers a number of advantages over competing measures of evidential confirmation that have been noted by others, and this all seems to reinforce Roush’s tracking approach to knowledge.

Of course, the LR for some e and h could be high without its being the case that $p(h/e)$ is high. This could be the case if the prior probability of h were suitably low. Now Roush’s intuition is that the question whether e is evidence for h is not independent of the value of $p(h/e)$. For she claims that e is ‘some evidence’ for h only if $p(h/e) > 0.5$, and e is ‘good evidence’ for h only if $p(h/e)$ is significantly higher than this. So the question of the extent to which e is evidence for h somehow must involve both facts about the LR and facts which determine a posterior for h that is not too low.

All of this is spelled out nicely in Chapter 5 of Roush's book, entitled "What is Evidence? Discrimination, Indication, and Leverage". The reference to 'leverage' pertains to a criterion for measures of evidential relevance that Roush defends. One way of moving a very heavy object is to place it beneath one end of a lever which is placed on a fulcrum – assuming the other end of the lever is farther away from the fulcrum than the heavy object, one can move the heavy object by placing a lighter object on the other end. The ability to move a heavier object by moving only a lighter object is obviously what makes this a useful technique. Similarly, one way to establish the truth or falsehood of some hypothesis is to acquire evidence that bears on that hypothesis – but for this method to be of any practical use it had better be the case that it is easier to establish the truth value of the evidence than to establish the truth value of the hypothesis. This relative ease of evidence determination constitutes Roush's intuitive picture of evidential leverage.

A piece of evidence *e*, even if it tracked the truth about *h*, would lack the virtue of leverage if the only way to establish the truth value of *e* required finding out whether *h* was true – as Roush reasonably points out. But Roush takes the concept of leverage still further: *e* lacks leverage, she explains, to the extent that using *e* as evidence for *h* (specifically, using *e* to push the posterior of *h* to a reasonably high level) requires independent information about the independent plausibility of *h*. *Ceteris paribus*, then, a method by which we use knowledge of *e* to compute a probability for *h* is better the less independent information about the prior probability of *h* that method requires. This is because such methods offer more 'leverage' than methods that require more of such information (Roush 2005 159).

But wait – I can hear some people saying – isn't this taking the concept of leverage a bit too far? For this seems perilously close to recommending methods of computing posterior probabilities for h that do not depend on solid information about the prior probability of h – and that sounds like it recommends methods of posterior computation that commit (dare I say it?) the base-rate fallacy. The base rate fallacy is, to put it as simply as possible, the fallacy of thinking that one can establish the probability of h given some evidence e without information about the prior probability of h .

Needless to say, Roush knows all about the base-rate fallacy, and she doesn't really intend to recommend a method of computing posteriors that literally ignores the prior probability of the hypothesis at stake. But the method she recommends nonetheless purports to respect the intuitive concept of leverage in the following way: the value of $p(h/e)$ can be computed, given facts about the probabilities of likelihood terms $p(e/h)$ and $p(e/\sim h)$, without *direct* information about the prior probability of h . Another type of information can do the work that information about the priors is usually supposed to do. This is information about the probability of e , $p(e)$. Given information about the value of $p(e)$, and the probabilities of the likelihoods, we can compute the value of $p(h)$ by starting with the theorem of total probability:

$$\text{(Total Probability)} \quad p(e) = p(e/h)p(h) + p(e/\sim h)[1-p(h)].$$

Clearly, given values for $p(e)$, $p(e/h)$, and $p(e/\sim h)$, we can solve for $p(h)$, by re-arranging the theorem of Total Probability – I call the resulting equation the Leverage Equation:

$$\text{(Leverage Equation)} \quad p(h) = [p(e) - p(e/\sim h)]/[p(e/h) - p(e/\sim h)]$$

So now we can add the value of $p(h)$ to the other values, and we have everything we need to compute $p(h/e)$ – we can just use Bayes' theorem! So it appears that a method that

computes $p(h/e)$ using independent information about $p(e)$, $p(e/h)$, and $p(e/\sim h)$ (but no “direct information” about $p(h)$) should be both possible and attractive because it respects the leverage ideal. I’ll call it the ‘Leverage Method’ of computing posterior probabilities.

How exactly is the Leverage Method supposed to work? Roush shows that the posterior probability $p(h/e)$ can be shown to be a function of the Likelihood Ratio, $P(e/h)$, and $P(e)$ – and that by plugging in values for these terms one can compute the value of the posterior probability (166-171). Rather than focus on these computations, however, I will focus on the example Roush uses to illustrate the leveraging method – an example she describes as a “simple, paradigm case.” (171-2) Suppose there is a medical test whose purpose is to determine whether a subject has a particular disease D . A blood test purports to determine whether the subject has a blood marker d for D . Let e = Mary tests positive for d and h = Mary has disease D ; Roush assumes that $p(e/h) = 0.95$ and $p(e/\sim h) = 0.15$ – the likelihood values reflect the probabilities of the subjunctive conditional statements “If h (or not h) were true, then e would be true.” That is, a subject is 95% likely to test positive if she has disease D , and only 15% likely to test positive if she doesn’t have D . Roush imagines that we do not know the prior probability of h – e.g., we do not know the relative frequency of D among the relevant population of which Mary is a member. But Roush claims that we can estimate the value of $p(h/e)$ nonetheless, if we are possessed of information about the value of $p(e)$. This is because information about the value of $p(e)$, together with the values of $p(e/h)$ and $p(e/\sim h)$, fix a value for $p(h)$ by way of the Leveraging Equation – as noted above to be just a re-arrangement of the Theorem of Total Probability. Once the value of $p(h)$ has been fixed in this way, it can be

plugged into Bayes' theorem to compute the value of $p(h/e)$, since the other terms on the right hand side of Bayes' theorem are now all apparently available:

$$\text{(Bayes' theorem)} \quad p(h/e) = p(h)p(e/h)/p(e)$$

But how is the value of $p(e)$ to be computed? Roush imagines that it can be fixed just by observations that – to one extent or another - directly confirm e . She offers a somewhat complex example of how this can be done in the Mary case: Roush imagines that Mary is tested ten times for d , and she tests positive for 8 out of 10 cases. This would entail, Roush claims, that $p(e) = 0.8$; $p(e)$ will thus be lower if she tests positive in fewer of the 10 cases, etc. I'm not quite sure what to say about this claim – I thought that being a positive tester was just a matter of producing a positive test outcome, so Mary (in the case described) was a positive tester 8 times and wasn't 2 times. I'm not sure that amounts to a case which should be described as one in which $p(e) = 0.8$. I gather the kind of example Roush is generating is one that provides a certain amount of observational evidence that confirms e but simply drives it to 0.8 instead of, say, confirming it conclusively. I hope it won't rankle if I propose a slightly different way of accumulating observational evidence for e that has the, I think, the intended effect: suppose Mary is tested a single time, and there is a report on which the outcome that she tested positive is written. But we are reading the report in a dark room and it is hard to see. We attempt to read the report 'by candlelight' (to use Richard Jeffrey's phrase) – and the partial illumination causes us to declare that $p(e) = 0.8$ (or less if the candle is less bright, etc.) Given that this way of imagining the example delivers the intended result about $p(e)$, I assume it will work about as well as Roush's examples involving multiple tests.

Roush then notes that if $p(e)$ is determined by observation to be 0.8, then this procedure determines that $p(h/e) = 0.96$. If on the other hand $p(e)$ is determined to be 0.2, then $p(h/e) = 0.3$. She gets these values in the way described: use the Leveraging Equation to derive $p(h)$ from $p(e)$ and the likelihood terms, and plug that value into Bayes' theorem along with the value of observationally fixed value of $p(e)$ and $p(e/h)$. And oila: the posterior of h on e has been determined without direct information about the prior probability of h , and the leveraging ideal has been fully respected.

But just how content should we be with these computations? I confess I am troubled. One rather odd feature of the example just described that bears noting is that the posterior probability of h goes up as we acquire observational evidence that confirms e . We are encouraged to make sense of this by noting that this method uses information about $p(e)$ to attempt to compute the prior probability of h – so a higher value for $p(e)$ supposedly indicates a higher value for the prior of h - and of course as the prior probability of some hypothesis goes up its posterior may well go up too, keeping the likelihood values fixed. But the elevating probability of h is supposedly just a reflection of the elevating probability of e . It isn't normally the case, where e confirms h , that the value of $p(h/e)$ goes up as the value of $p(e)$ is driven up by gathering observations that directly confirm e . Rather, in the typical case, as these observations are gathered and $p(e)$ inches up, the probability of h will inch up toward a fixed posterior $p(h/e)$. If you are considering the hypothesis that I will die of lung cancer, your probability may go up as you acquire observations that suggest I am a heavy smoker. First you note that characteristic tobacco smell on my clothes, then it's the yellow of my teeth, and up goes the probability of the evidence that I smoke heavily bit by bit. This will push up the

probability that I will die of lung cancer bit by bit – but not the posterior probability that I will die of lung cancer on the assumption that I am a heavy smoker. But this latter pushing seems entailed by Roush's leveraging method. At least it's a bit puzzling.

Let me tell the story of Mary with a twist. Keep everything the same, but suppose that before we compute the value of $p(h/e)$ we become aware of what would normally be called the prior probability of h . Let's suppose that we find out, e.g., that among the relevant population of which Mary is a member, the relative frequency of disease D is very low – say 0.00001. So we set $p(h) = 0.00001$. Now, we want to determine $p(h/e)$. Now what value should we use for $p(e)$? Roush considered a case in which $p(e)=0.8$ because some observations had pushed up the probability of e – but notice that our decision to set $p(h) = 0.00001$ means that my probability function 'p' is computed without taking into account additional observational evidence that might confirm h (such as observations that confirm e). This means that $p(e)$ will be the probability that e should be assigned prior to the acquisition of any such observational evidence. This, after all, would seem to coincide with the intuitive Bayesian notion of the 'prior probability of e '. Now $p(e) = p(e/h)p(h) + p(e/\sim h)[1-p(h)]$ – but we have values for all these terms, and they entail that $p(e) = (0.95)(.00001) + (0.15)(.99999) = 0.1500084$. But now Bayes' theorem gives $p(h/e) = 0.0000633$.

Now we have a conundrum – the additional information about the prior probability of h that I supposed in making this computation has produced a value for $p(h/e)$ that is extremely low – if e were to be shown true (i.e. if Mary tested positive for d) she could still rest easy knowing that e was almost certainly a false positive – simply because, given the assumed probabilities, it is much more likely to be a false positive

than for it to be the case that she has the disease. This value isn't any thing like the values for $p(h/e)$ that Roush computes using the leveraging method – remember that, assuming $p(e) = 0.8$, the value of $p(h/e)$ is 0.96! Of course the leveraging method gives much lower values for $p(h/e)$ as $p(e)$ drops, but I've noted already that there is something amiss when posterior probabilities rise or fall based on whatever observational evidence happens to exist that supports e .

Of course Roush might object that this is no conundrum. I computed the value of $p(h/e)$ using a value for $p(h)$ that she did not assume, and using a value for $p(e)$ that was consequently very different from hers. Let's pause at this point and consider what probability should be assigned to h – on my assumption about the low prior of h - when it comes to be the case that we acquire the candlelight observation that drives the value of $p(e)$ up to 0.8 – we can do that by Jeffrey conditionalization. But since the value of $p(h/e)$ is extremely low, and e is evidence for h , it follows that $p(h/p(e)=0.8)$ will be even lower. But this super low value for the probability of h , given that $p(e) = 0.8$, is nothing like the value delivered by the Leveraging Equation as deployed by Roush (which is 0.8125!) But if the leveraging method were a sound method of computing the value of h 's probability given some observationally fixed input value for e 's probability, why would it matter if we go ahead and make some explicit assumption about the prior probability of h ? Shouldn't the Leveraging Method deliver the same result for $p(h/e)$ as ordinary Bayesian computations of $p(h/e)$?

So I say there is a conundrum here after all. The sticking point, as I see it, concerns how it is that the leveraging method produces the curious values for $p(h/e)$ that it does. Curious, in part, because it makes the value of $p(h/e)$ critically dependent on

observationally fixed values of $p(e)$. What is wrong with the leveraging method? In my opinion, the leveraging method suffers two problems as it is used by Roush in the Mary example. One of these is that the method as used by Roush commits the base rate fallacy after all – the other is that it suffers a problem that I deem the ‘fixed likelihood fallacy’. I deal with each in turn. After I have argued for these points, I will go on to explain how I think the Leveraging Method could be used as a valid method.

Roush imagines that as we acquire observations that support e (Mary is a positive tester) we fix the value of $p(e)$ just on the basis of those observations. So, on my version of the case, when we acquire the candlelight observation that by itself renders it 80% likely that e is true, we set $p(e) = 0.8$ (or, on her version, when Mary tests positive 8 out of 10 times). But wait – prior to the observation we performed – what was the prior probability of e ? Let me deem the probability function that we used prior to the candlelight observation (which I refer to as ‘ C ’) as p^* . So $p^*(e/C) = p(e)$. $p^*(e/C)$ could be significantly different from 0.8 – say if $p^*(e)$ is very low or very high. (And of course $p^*(e)$ will presumably be very low if $p^*(h)$ is very low). For of course by Bayes’ theorem $p^*(e/C) = p^*(e)p^*(C/e)/p^*(C)$. By ignoring the relevance of $p^*(e)$ in computing $p^*(e/C)$ the Leveraging Method essentially commits the base rate fallacy – it ignores the relevance of what might be a crucial prior probability.

Of course Roush might object that it is common practice in Bayesian thought to fix $p(e)$ just by observation – say in cases where observation shows that e is conclusively (or nearly conclusively) true. I agree that in cases like these the relevance of $p^*(e)$ can be drowned out by the overwhelming probative weight of observation. But Roush seems interested in cases, such as the Mary case, where observation is less conclusive – where it

simply fixes some intermediate value for the probability of e. In cases like these, the prior probability of e could be quite important.

How should the prior probability of e, $p^*(e)$, be computed? Quite possibly by the Theorem of Total Probability applied to p^* :

$$p^*(e) = p^*(e/h)p^*(h) + p^*(e/\sim h)[1-p^*(h)]$$

But of course to use this method we would have to know the value of $p^*(h)$, the prior probability of h, and that is a value that the Leveraging Method purports to do without. So insofar as the Leveraging Method needs a value for $p^*(e)$ to compute the value of $p(e)$, it apparently needs a value for $p^*(h)$ – by offering a value for $p(h/e)$ which has no access to $p^*(h)$, the Leveraging Method can be seen to commit the base rate fallacy in another way. That, as I see it, is the Bayesian law of the land: no priors means no posteriors.

From a certain angle, it seems to me somewhat clear that Roush's Leveraging Method could not be expected to deliver a reasonable value for the probability that Mary is sick with disease D. For it purports to compute this probability on the basis of too little information: given just an observationally fixed value of $p(e)$, which of course is just observational evidence that she was a positive tester, and the probabilities of the likelihood terms, the value of $p(h/e)$ is computed. But wait – how are values for these terms supposed to tell us how likely it is that Mary is sick given that she tested positive – in the complete absence of information about how common or rare the disease is in general (i.e. absent information about $p^*(h)$)? Such posterior probabilities are notoriously sensitive to varying information about such relative frequencies. Of course, the idea of the leveraging method was to tease out information about the prior of h given

a certain value for the prior of e – but it just can't be that genuine information about the probability of h can be 'teased out' of information about the observationally fixed value of $p(e)$ and the original likelihood values. The information isn't there to be found.

The second fallacy that I believe the Leveraging Method as deployed in the Mary case commits I call the fixed likelihood fallacy. Recall that Roush sets the value of $p(e/h) = 0.95$, which reflects the fact that 95% of people with disease D test positive, and $p(e/\sim h) = 0.15$, which reflects the fact that 15% of people without D test positive. When the new observational evidence comes in that confirms e , she uses the same values for these likelihood terms in the Leveraging Equation to compute a value for $p(h)$. But this ignores the fact that incoming observational evidence (such as C , the candlelight observation) that confirms e could dramatically alter the values of the likelihood terms. To see how this takes place, consider the following instance of Total Probability:

$$p^*(e/C) = p^*(e/h\&C)p^*(h/C) + p^*(e/\sim h\&C)p^*(\sim h/C)^1$$

Once C is possessed, in other words, the likelihood terms can be expected to change considerably. The value of $p(e/h)$ is equal to $p^*(e/h\&C)$, but this will establish a probability for e conditional both on the assumption that h is true and the candlelight observation, which offer in effect two distinguishable reasons to believe that e is true. Similarly for the value of $p^*(e/\sim h\&C)$: given the evidence C offers for e , this value could be well above the value of 0.15 that Roush assigns to it. By keeping the values of the likelihood terms fixed as evidence for e comes in, Roush ignores the effect of that evidence on the likelihood probabilities – I call that the fixed likelihood fallacy.

Roush might reply that while the values of the likelihoods may change, the likelihood ratio LR could remain the same. While this is in some sense a theoretical

¹ I am grateful to Patrick Maher for bringing this relevant equation to my attention.

possibility, I see no reason to think that it will be true in general. In the Mary case, for example, it seems to me that the likelihood terms could be brought quite close to each other given the ability of the candlelight observation to substantially increase the probability of e assuming $\sim h$. The imposition of a requirement that the updated values of the likelihood terms be consistent with a fixed LR strikes me as arbitrary.

All this makes it sound like I think the Leveraging Method is completely wrong. But I don't. In fact, I think it's an interesting and perfectly valid method – under certain restrictions. Here's the way I would use it: let's go back to the Mary problem one more time. Suppose Mary is a member of a remotely located population of the Nazika people (I just made the name 'Nazika' up). We have no observational evidence that confirms that Mary is a positive tester as yet, so our probability function is p^* . We don't know the prior probability that Mary has disease D, for we don't know the relative frequency of D among the Nazika people. However, we do know the value of the likelihood terms $p^*(e/h)$ and $p^*(e/\sim h)$ from our study of the test on other populations (that is, the probability that a human subject will test positive given that she does, or does not, have the disease) and there is no reason not to apply these to the Nazika people also. Now suppose we want to know the posterior probability that Mary has D given that e is true (i.e. given that she tested positive). For this we need a respectable candidate for the $p^*(h)$ – which as yet we don't have. But suppose we have done the test on a large number of randomly sampled members of the Nazika people, and have determined the relative frequency of Nazika people who test positive. I say use this relative frequency to fix the value of $p^*(e)$, plug this value into the theorem of total probability and compute $p^*(h)$ using the established values of the likelihood terms. The resulting value for $p^*(h)$

should measure the relative frequency of D among the Nazika people, and that I will call the prior probability of h – the actual value of $p^*(h)$. This is because this value of $p^*(e)$ isn't the kind of thing I've been calling an observationally based value, i.e., isn't based on the results of having applied the test to Mary (either several times or once). Rather, it is in what I am inclined to call in a certain pure sense a prior probability for e – the probability that we would assign to e in the absence of any observational evidence that directly supports e. I'll call $p^*(e)$ the 'pure prior probability' of e. We expect the relative frequency of positive testers will simply consist of positive testers who have the disease and false positive testers who don't – but given the well established values of $p^*(e/h)$ and $p^*(e/\sim h)$ this means that this relative frequency will almost certainly fall between 0.95 and 0.15. (We can use these values of the likelihood terms in good conscience because they have not been modified by observational evidence for e at this point). Thus this relative frequency should provide an indication of the actual relative frequency of disease D among the Nazika . Plug this prior probability of h, i.e. $p^*(h)$, along with the described value of $p^*(e)$, along with the likelihood terms, into Bayes' theorem and compute the probability that Mary has D given that she tested positive, i.e. $p^*(h/e)$. Not only is this formulation of the leverage method perfectly valid, insofar as it avoids both the base rate fallacy and the fixed likelihood fallacy, it seems to offer precisely the advantage that Roush cited: it enables us to compute the posterior probability of h without direct information about the prior plausibility of h – but only by using information about the pure prior probability of e and teasing out information about the associated prior probability of h. It seems to me to be the case that this kind of leveraging technique could be useful and valid in certain types of cases. Cases, that is, in which for some

reason we have easier epistemic access to the value of the prior of e than we do to the prior of h.

Prior to presenting the Mary example as an illustration of the leveraging method, Roush presents a number of results in Chapter 5 that are based on variations of this method. She shows, e.g., that by re-arranging Bayes' theorem we have:

$$P(h/e) = [LR - P(e/h)/P(e)]/[LR - 1] \quad (\text{Roush 166})$$

The way in which values for $P(h/e)$ are determined by the terms on the right hand side are portrayed in three dimensional diagrams that show how particular likelihood ratios and particular values for $p(e/h)$ can uniquely determine values for $p(h/e)$ given fixed values for $p(e)$. It seems to me that the derivation of this equation depends upon the same basic move that I claimed made for possible trouble in the Mary example: the value of $p(h)$ is derived from the theorem of total probability (see Chapter 5 footnote 11). If the value of $p(e)$ is based on mere observation alone, then its establishment could be accused of committing the base rate fallacy, etc. If the value of $p(e/h)$ (which occurs in the above equation) is kept at a fixed value that ignores observational evidence for e, then we seem to have the fixed likelihood fallacy. But it is not inevitable that either of these fallacies are being committed. My position is that the equation above along with all the three dimensional diagrams are fine, provided that the value for $p(e)$ is equated just with the pure prior probability for e – not some observationally fixed value for this probability. And throughout this earlier discussion Roush says nothing, I believe, that prevents her or us from interpreting $p(e)$ as the pure prior probability of e – she needn't revise a line of this earlier portion of the chapter. It is only the Mary example which commits anything I would call an error by suggesting that the value of $p(h)$ can be computed by way of the

leveraging method when $p(e)$ is fixed just observationally and the likelihood terms are not updated appropriately.

The distinction between prior and posterior probabilities has never been a super clear one, but I think we ignore it at our peril. To make this point again I would offer a few remarks on the latter portion of Chapter 5. Roush notes that “the usual Bayesian story” is that e confirms h better the lower the value of $p(e)$ (172). Roush takes a different view, given her commitment to the leveraging method, since the gist of this method is to use the value of $p(e)$ to compute, indirectly, $p(h)$ – so high values of $p(e)$ can in principle push up the prior, and thus the posterior, probability of h . So she defends the preference for high $p(e)$ over the standard preference for low $p(e)$ at some length. For example, she emphasizes that one advantage of high $p(e)$ is that a high value amounts to strong evidence that e is true – and surely, if e is to be evidence for h , we need such evidence. This comment, however, invites the quick rejoinder that the old Bayesian preference for a low value of $p(e)$ was not a preference for a low value of the posterior probability of e , which is of course what we need if we are to actually use e as evidence for h in the act of conditionalization. It was supposed to be a preference for a low prior probability for e – an entirely different preference altogether. But of course Roush has, on behalf of the high prior probability of e preference, another point to make – and that is that a high prior for e can – in some cases - be correlated with a high value for the prior probability of h – and of course high priors for h can make for high posteriors. This point – which I have argued at some length is perfectly valid if by $p(e)$ we mean the pure prior probability of e – is interesting and important, and it does count against the old Bayesian slogan. But a preference for a high prior probability for e is not to be confused with a

preference for a high posterior probability for e . By simply talking about whether the value of ' $p(e)$ ' should be high or low, this distinction in preferences is obfuscated somewhat in Roush's discussion. The same obfuscation pervades, in a somewhat different way, Roush's presentation of the Mary example. Let us keep our priors and posteriors separate.

Reference

Roush, Sherrilyn. (2005), *Tracking Truth: Knowledge, Evidence and Science* (New York and Oxford: Oxford University Press).