ARE THERE BAYESIAN SUCCESS STORIES?: THE CASE OF THE RAVENS PARADOX

Michael Kruse

Department of Philosophy

Virginia Polytechnic Institute and State University

ARE THERE BAYESIAN SUCCESS STORIES?: THE CASE OF THE RAVENS PARADOX[†]

ABSTRACT. An oft-cited virtue of subjective Bayesian confirmation theory is its resolution to Hempel's "Ravens Paradox". I consider a variation on the Ravens Paradox that seems to resist a principled Bayesian resolution, and discuss a non-Bayesian analysis that allows us to distinguish the effect of data on inferences in terms of how they are generated. I then draw out some implications of this result for the prospects of Bayesianism as a general philosophy of science.

1. The Ravens Paradox.

The so-called "Ravens Paradox" arises from the intuition that not all observations that are *logically consistent with* a hypothesis constitute the same *evidence* for that hypothesis. An observation of a black raven, $Ra\&Ba$, is intuitively taken as better evidence for the hypothesis h: $(\forall x)(Rx \supset Bx)$ than an observation of a non-black non-raven, $\sim Rb\&\sim Ba$. But if h is confirmed by $Ra\&Ba$ by the *logical* relationship between the two, then $\sim Rb\&\sim Bb$ confirms the logically equivalent claim that $(\forall x)(\sim Bx \supset \sim Rx)$. Since evidence for h is evidence for any logically equivalent h', we have the 'paradoxical' result that $\sim Rb\&\sim Bb$ confirms h.

Subjective Bayesian confirmation theory offers a resolution to this paradox.[1] For Bayesians, the *evidential import* of e on h — basically, the effect of e on inferences about h — depends on how much e confirms h, *i.e.*, the difference between $P(h)$ and $P(h/e)$.[2] The Bayesian resolution turns on the fact that under plausible conditions, the posterior probability of h given $Ra\&Ba$ — $P(h/Ra\&Ba)$ — is *greater* than $P(h/\sim Rb\&\sim Bb)$.[3]

[†] Thanks to Branden Fitelson for his very useful comments on an earlier draft of this paper.

[1] My discussion is confined to subjective versions of Bayesianism of the sort presented in Earman (1992) and Howson and Urbach (1992).

[2] See Fitelson (1998) for an excellent discussion of the variety of methods of measuring increases in probability and the implications of this for Bayesianism.

[3] Note that the Bayesian's aim is *not* to show that $Ra\&Ba$ *always* confirms the hypothesis that all ravens are black more than $\sim Ra\&\sim Ba$, but rather to identify when it does.

This ability to discriminate between the *evidential import* of these observations is taken to show how Bayesianism helps to justify methodological judgments in science. I shall argue, however, that this "success story" is less impressive than it might first appear. While Bayesianism *does* allow us to distinguish the evidential import of *certain* sets of data, it *obscures* distinctions between *other* sets of data — distinctions that seem *just* as relevant as that between observations of black ravens and white shoes. This, in turn, reveals a problem for Bayesianism's ability to justify basic judgments concerning scientific evidence.

2. The Bayesian Resolution.

The Ravens Paradox arises when we compare the effect of two different observations on inferences about hypotheses concerning the proportion of ravens that are black. In a statistical approach like Bayesianism, we represent these observations with random variables: X is the number of black things found in an n-fold sample from the set of ravens, $\boldsymbol{R}$, and Y is the number of ravens found in an n-fold sample from the set of non-black things, not-$\boldsymbol{B}$. Hypotheses about the proportion of ravens that are black are represented by parameter values $0 \leq \theta \leq 1$, with $\theta = 1$ corresponding to the claim that all ravens are black.

The Bayesian resolution shows that $P(\theta = 1 / X = n) \geq P(\theta = 1 / Y = n)$, where $Y = n$ corresponds to seeing n non-black non-ravens and $X = n$ corresponds to seeing n black ravens. The two conditions below are jointly sufficient to show this result:

(A) $P(B) = P(B/h) = P(B/h') = P(B/h')$ and $P(R) = P(R/h) = P(R/h') = P(R/h')$ for any two hypotheses h and h': the probability of a randomly selected individual being black (or being a raven) is *independent* of the probability of a randomly selected raven being black.

(B) $P(R) < 1 - P(B)$: there are fewer ravens than non-black things in the population.

Given these assumptions, Bayesians can show that $X=n$ is *better evidence* for $\theta=1$ than $Y=n$.[4]

The key to the Bayesian resolution lies in the *likelihood functions* defined with respect to these data. The likelihood function is a function of the parameter of interest — in this case, θ — which defines the probability of the data under each possible value of θ. If we observe $X=x$ black ravens in n samples from $\boldsymbol{R}$, the likelihood function, $L_X(\theta)$, is:

$$L_X(\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

To find $L_Y(\theta)$, first note that Y is an estimate of the proportion of members of not-$\boldsymbol{B}$ that are also members of not-$\boldsymbol{R}$; call this proportion ϕ. So, $L_Y(\phi)$ is:

$$L_Y(\phi)=\binom{n}{y}\phi^y(1-\phi)^{n-y}$$

Given condition (A) above, Bayes's Theorem implies that:

$$\phi=1-\frac{P(R)}{P(not-B)}(1-\theta)$$

Making substitutions yields:

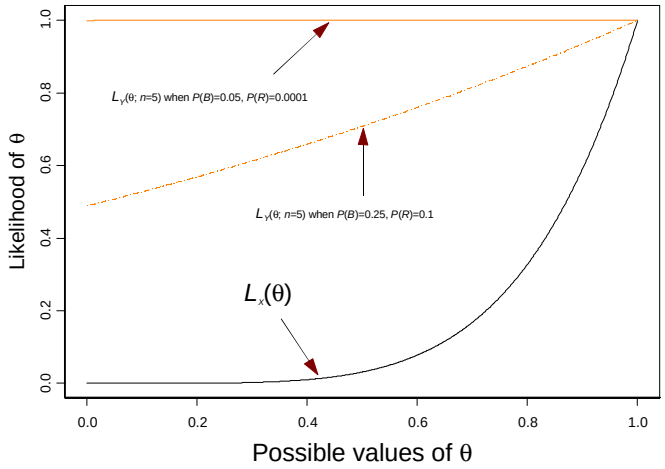
$$L_Y(\theta)=\binom{n}{y}\left(1-\frac{P(R)}{P(not-B)}(1-\theta)\right)^y\left(\frac{P(R)}{P(not-B)}(1-\theta)\right)^{n-y}$$

When $X=Y=n$, both $L_X(\theta)$ and $L_Y(\theta)$ are maximized when $\theta=1$, *i.e.*, the hypothesis that *fits* each observation best is $\theta=1$. As the figure below indicates, for every *other* $\theta<1$, $L_Y(\theta)>L_X(\theta)$ — X does a better job of *discriminating* among the values of θ than Y.[5]

---

[4] See, for instance, Earman (1992), Horwich (1982, 1993), Hosiasson-Lindenbaum (1940) and Howson and Urbach (1992) for variations on this Bayesian account.

[5] In the figure below, $X=Y=5$ and $n=5$. $L_X(\theta)$ is independent of $P(B)$ and $P(R)$. $L_Y(\theta)$ is plotted for two cases: $P(B)=0.05$ and $P(R)=0.0001$ (solid orange line) and $P(B)=0.25$ and $P(R)=0.1$ (dashed orange line). Given conditions (A) and (B) above, $L_Y(\theta)\geq L_X(\theta)$ for all θ.



Given this difference between $L_X(\theta)$ and $L_Y(\theta)$, the Bayesian resolution is straightforward. By Bayes's Theorem, the posterior probability of $\theta=1$ is inversely proportional to the probability of the data, which is an average of the likelihoods weighted by the prior probabilities of the values of θ. Since $L_Y(\theta)>L_X(\theta)$ when $\theta<1$ and $L_Y(\theta)=L_X(\theta)$ when $\theta=1$, $P(Y=n)>P(X=n)$, and so, $P(\theta=1/Y=n)\leq P(\theta=1/X=n)$.

The difference in the 'boost' that each observation gives the hypothesis is due to the difference between $L_X(\theta)$ and $L_Y(\theta)$. Indeed, this is the *only* factor that can distinguish the evidential import of two different observations for a Bayesian, since data enter into Bayes's Theorem only by way of the likelihood function.[6]

If differences between the likelihood functions really do correspond to intuitive differences in evidential import, Bayesianism has scored an impressive success. But does this correspondence hold? As a test, consider a *third* sampling procedure in which we draw

[6] Since the evidential difference between these observations depends only on the likelihoods, one can distinguish these observations without embracing Bayesianism. (See Royall 1997.) The uniquely Bayesian feature of the account above comes from using the difference in likelihood functions to calculate posterior probabilities.

from ***B*** until we find n ravens, and let $Z=z$ be the number of samples needed to find the n^{th} raven. The proportion n/z gives us an estimate of $\tau=P(R/B)$. In this case, the likelihood function defined with respect to Z is:

$$L_Z(\tau)=\binom{z-1}{n-1}\tau^n(1-\tau)^{z-n}$$

Again, given condition (A), Bayes's Theorem implies that $\tau=k\theta$, where $k=P(R)/P(B)$. Substituting this for τ yields:

$$L_Z(\theta)=\binom{z-1}{n-1}(k\theta)^n(1-k\theta)^{z-n}$$

In selecting an experiment to run, it is sensible to prefer sampling n times from ***R*** rather than sampling from ***B*** until we find n ravens. This judgment has a Bayesian rationale. If, for instance, $\theta=1$, $n=5$ and $k=0.002$, we *expect* the observed value of Z to be around 2500. Assuming a uniform prior for θ, the expected posterior density of $\theta=1$ would be $P(\theta=1/X=5, n=5)=6$, while $P(\theta=1/Z=2500, n=5)\approx 2.29$. So, if $\theta=1$, we *expect* that sampling from ***R*** will yield a posterior for $\theta=1$ greater than if we were to sample from ***B***, *i.e.*, sampling from ***R*** gives us a better chance of highly confirming $\theta=1$ than sampling from ***B***.

This is a justification for running one experiment rather than another — it justifies a *pre-trial* decision to observe X rather than Z. But what about *post-trial* assessments of evidence? For instance, say we are given five black ravens and are told they were found *either* by sampling five times from ***R*** *or* by sampling from ***B*** until finding five ravens (which, surprisingly enough, were found in the first five samples). Would learning how these data were generated make a difference to our inference?

Given condition (A), the Bayesian must *deny* any such difference. To see why, note that for $n=5$ and $X=Z=5$, $L_X(\theta)$ and $L_Z(\theta)$ are:

$$L_X(\theta) = \binom{5}{5}\theta^5 = \theta^5$$
$$L_Z(\theta) = \binom{4}{4}(k\theta)^5 = k^5\theta^5$$

$L_X(\theta)$ and $L_Z(\theta)$ differ only by a constant — $L_Z(\theta)=k^5L_X(\theta)$. So, no matter how the data were generated, our posterior distribution will be the same. With respect to the *evidential import* of these two results, there can be no difference for a Bayesian — even though the second method guarantees that any raven observed would *have* to be black.

Surely, this observation of Z tells us *something* about θ. As we will see in section 4, however, there are Bayesians who agree that there is *some* difference between the evidence that $X=Z=5$ provide for $\theta=1$. But if there *is* an intuitive difference in what these outcomes tell us about θ that the Bayesians cannot capture (again, when condition (A) obtains), it seems we have another version of the Ravens Paradox. Before, a *logical* measure of support couldn't distinguish observations with (intuitively) different evidential meaning; here, a *probabilistic* measure suffers the same defect.

3. An Alternative Account.

In section 4 below, I consider how Bayesians have responded to this result. First, however, it is worth considering how we might justify the intuitive difference in the evidential import of observing $X=Z=5$. To motivate this *non-Bayesian* account, it is useful to think of what we are trying to do when we gather data.

One goal in science is to use data to select *accurate* parameter values (*i.e.*, hypotheses), ones that will yield predictions that have a good fit with new data.[7] The statistical 'device' used to select a parameter is an *estimator*, *i.e.*, a function of the observed variable that returns a parameter value as an estimate. What we want, then, is an estimator with properties that helps to justify the claim that the estimate it returns is close to the *true* value of θ, θ^*.

This concerns the *reliability* of our estimator. While there may be various ways of characterizing the reliability of an estimator, two properties seem central. One is the estimator's *sensitivity* to the truth: do the estimates we make of θ^* depend on what the *actual* value of θ^* is? One way for this condition to be satisfied is if the estimator is *unbiased*, in which case its mean value *is* the truth.[8]

Another relevant property is the estimator's *variance*, which indicates how closely concentrated estimates fall around the estimator's mean value. In general, given a *biased* and an *unbiased* estimator, an assessment of relative reliability demands considering *how* biased the first is and the *differences* in their variances. (For instance, a slightly biased estimator with a small variance may be more reliable than an unbiased estimator with a large variance.)

An important special case that in which we compare *unbiased* estimators; here, the one with the *smallest* variance is the most reliable one, since it is most likely to yield an estimate that is close to the truth. It happens that this special case allows us to distinguish the reliability of the estimators used in the three sampling procedures discussed above.

---

[7] This is essentially the notion of predictive accuracy as defined in Foster and Sober (1994).
[8] Unbiasedness is not *necessary* for reliability in this sense, since an estimator may be biased, yet still deliver estimates concentrated closely around the truth.

In these three sampling procedures, one important method of estimation — the method of maximum likelihood — implies three different estimators of θ^*, depending on which variable we are observing:[9]

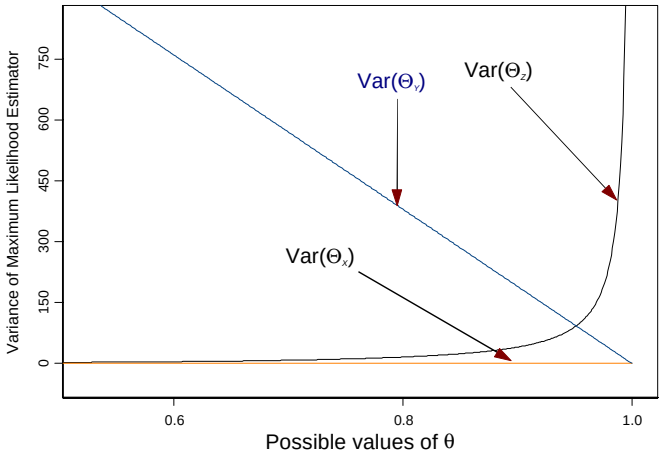
$$\Theta_X = \frac{X}{n}$$
$$\Theta_Y = 1 - \left(\frac{1-P(B)}{P(R)}\right)\left(1-\frac{Y}{n}\right)$$
$$\Theta_Z = \left(\frac{P(B)}{P(R)}\right)\left(\frac{n}{Z}\right)$$

Each of these estimators is a function of a random variable, and so each is a statistic with a mean and a variance. As maximum likelihood estimators, each is asymptotically unbiased, *i.e.*, as $n \to \infty$, the mean value of each converges on θ^*. Their variances are given below.

$$\text{var}(\Theta_X) = \frac{\theta^*(1-\theta^*)}{n}$$
$$\text{var}(\Theta_Y) = \left[\frac{1-\theta^*}{n}\right]\left[\frac{1-P(B)}{P(R)}\right]\left[1-\frac{1-P(B)}{P(R)}(1-\theta^*)\right]$$
$$\text{var}(\Theta_Z) = n\left(\frac{\theta^{*2}}{1-\theta^*}\right)$$

As these indicate, the variance of each estimator depends on θ^*. The figure below shows the variances of each estimator over a range of values for θ when n=5, $P(B)$=0.05 and $P(R)$=0.0001. Note that if θ^* is between approximately 0.80 and 0.99, var(Θ_X) is significantly lower than either of the others.

[9] In Θ_X and Θ_Y, n is the number of samples; in Θ_Z, n is the number of ravens needed before we stop sampling.



So how does this relate to the Ravens Paradox? First, the differences between the variances of these three estimators give us good reason to prefer sampling from ***R*** rather than from either not-***B*** or ***B***.[10] This *pre-trial* judgment can be justified by noting that as long as θ is not extremely small, estimates from Θ_X will tend to fall nearer to the true value of θ than those from either Θ_Y or Θ_Z.[11] (Since we don't know θ^*, it is reasonable to use the estimator whose variance is at least as low as that of the others for plausible values of θ^*.)

My more controversial claim is that the differences between the estimators are also relevant *after* the data are in hand.[12] To motivate this claim, consider again the case of

---

[10] Forster (1995) has showed this for Θ_X and Θ_Y, and has demonstrated an important connection between the variance of estimators and the *predictive accuracy* of estimates, *i.e.*, the expected fit between a hypothesis and new data.

[11] $var(\Theta_X)>var(\Theta_Z)$ when θ^* is very small. (This fact helps to justify the methodological rule of conducting exponential experiments — a negative binomial experiment that stops after a single 'success' — when θ^* is assumed to be very small, *e.g.*, in estimating failure rates in computer systems.) In the ravens example, it is plausible to assume that θ^* is rather high, in which case we have reason to assume that $var(\Theta_X)<var(\Theta_Z)$.

[12] This amounts to rejecting what is known as the *Likelihood Principle*, an implication of Bayes's Theorem asserting that *all* information from an experiment relevant to inferences about a parameter is in the likelihood function. For arguments in favor of the Likelihood Principle, see Berger and Wolpert (1988), Edwards (1992) and Royall (1997).

observing five black ravens and being told that they came *either* from drawing five times from ***R*** *or* from sampling from ***B*** until finding five ravens. In *either* case, the estimate of θ^* is 1.

We have seen that if assumption (A) obtains, learning how those data were generated makes *no* difference to the likelihood functions, and so makes no difference in the *fit* between the data and any value of θ (as measured by the likelihood of θ). Further, since such information makes no difference for the likelihood functions, it can make *no* difference to a Bayesian's inferences about θ.

If we are interested in assigning *probabilities* to values of θ, this is surely correct. However, if, as seems plausible in scientific contexts, we want to find an *accurate* estimate of θ, the matter seems rather different. Given the *kind of inference* we want to make, learning how those data were generated *does* seem relevant in that the method of data-generation provides additional information relevant to inferences about the estimate's accuracy.

Why would this be relevant information? Imagine, for a moment, that that we draw a ball with a number printed on it from one of two urns. The balls in the first urn each have a number printed on them, and these numbers are narrowly distributed around θ^*, while balls in the second urn have numbers that are distributed more widely around θ^*. Clearly, if given the choice of drawing from one or the other urn, we would opt for the first — this goes to the *pre-trial* judgment. But it *also* seems that even *after* the ball has been drawn, knowing which urn it came from makes a difference to our inference about our 'estimate' of θ^*. In particular, if the ball came from the *first* urn, we have better justification for thinking it is close to θ^* than if it had come from the *second*.

This simple urn example illustrates how methods of data-generation can affect inferences about the accuracy of a hypothesis. As long as θ^* is not extremely small — which is

plausible enough in the raven example — estimates from Θ_X will tend to be closer to θ^* than an estimate from Θ_Z, just as draws from the first urn will tend to be closer to θ^* than draws from the second. So, as in the urn example, it seems that knowing how the data were generated *should* affect our inference about the accuracy of our estimate. In particular, while Θ_X and Θ_Z both yield the same estimate from the same set of data, we have reason to think that Θ_X's estimate is more accurate. This, in turn, helps to underwrite the intuition that the sample from ***R*** is better evidence for the accuracy of $\theta^*=1$ than the sample from not-***B***.[13]

Using differences between Θ_X and Θ_Z to distinguish what the data from these two experiments tell us about θ is analogous to discriminating between the same testimony offered by a *reliable* and an *unreliable* witness. In each case, the *observation* is clearly the same, yet the difference in how those observations came about seems to make an evidential difference, *i.e.*, they make a difference in what the data tell us about the quantity of interest.[14]

4. A Bayesian Response.

Above, I have argued that when assumption (A) obtains, Bayesians can draw no evidential difference between observing *n* black ravens drawn from ***R*** and *n* black ravens drawn from ***B***. The strength of the intuition that there *is* such a difference is suggested by the efforts of some to find a Bayesianism rationale for this intuition.[15]

---

[13] The claim here is that learning how the data were generated affects inferences about the *accuracy* of hypotheses, not their *probabilities*. To the extent that there *is* an intuitive difference between the evidential import of the observations of *X* and *Z*, this may suggest that concerns for *accuracy* drive scientific intuitions more than concerns for *probability*.

[14] The difference between basing assessments of evidential import on the likelihood function alone and claiming that evidential import depends on properties of an estimator or test is an instance of the contrast between what Mayo (1996) calls *evidential-relation* accounts and *error-statistical* accounts.

[15] See, for instance, Horwich (1993), Howson and Urbach (1992) and Korb (1994).

It is difficult to see how this could be done in any principled way. By Bayes's Theorem, after all, the posterior distribution of θ depends *solely* on the prior distribution of θ and the likelihood function, $L(\theta)$. If condition (A) holds, differences in how the data were obtained can make *no* difference to the Bayesian's inference.

Of course, if (A) does *not* hold, it may be that $P(\theta=1/X=n)\neq P(\theta=1/Z=n)$, since then $L(\theta; X=n)$ and $L(\theta; Z=n)$ need not be proportional.[16] So, just as when we compared sampling from ***R*** and not-***B*** in the original formulation of the Ravens Paradox, there *are* conditions on which our judgment that $X=n$ is better evidence for $\theta=1$ than $Z=n$ *coincides* with the fact that $P(\theta=1/X=n)\geq P(\theta=1/Z=n)$.[17]

But to turn these Bayesian accounts into *justifications* of these judgments, it is important to say why we should believe these conditions obtain. The mere fact that assumptions are required is *not* the problem.[18] However, it is worth noting that the assumptions sufficient for Bayesians to distinguish $X=n$ and $Y=n$ are ones for which we have justification *independent* of any commitment to Bayesianism — a fact that probably goes a long way toward explaining why the Bayesian resolution seems so successful. To explain why $X=n$ and $Z=n$ differ in their evidential import, however, we must assume that (A) fails — an assumption for which

---

[16] See Howson and Urbach (1992, 370-371). Korb (1994) *endorses* the intuition that the procedure used to generate the data *should* have an effect on confirmation. The example he gives shows this, but it is worth noting that it is one in which (A) does *not* hold, which allows for $P(\theta=1/X=n)>P(\theta=1/Z=n)$.

[17] Even then, it is not clear how this justifies the intuition that finding n black ravens is *weak* evidence for $\theta=1$ when we know *any* raven we see must be black. *That* judgment seems sound even when (A) holds, in which case $P(\theta=1/X=n)$ and $P(\theta=1/Z=n)$ *must* be equal.

[18] See note 11 above for how assumptions about the true value of θ can affect the account described in section 3.

there seems rather little justification beyond the fact that this is *needed* to make the Bayesian account work.[19]

The problem of motivating the conditions required by the Bayesian account is only compounded when we try to account for both the intuitive difference between $X=n$ and $Z=n$ *and* that between $X=n$ and $Y=n$. To distinguish $X=n$ and $Z=n$, (A) *must* fail. That, however, opens the door to examples like Good's (1967) in which $Y=n$ confirms $\theta=1$ more than $X=n$ does (despite the fact that ravens may *still* be less common than non-black things.) Conversely, assuming that (A) holds is a plausible way to show that $X=n$ is better evidence for $\theta=1$ than $Y=n$, yet doing so makes $X=n$ and $Z=n$ evidentially equivalent to the Bayesian.

This is not to say that one *couldn't* identify one set of conditions that match our intuitions in both cases. But what motivates the search for those conditions? If we make these assumptions *solely* to match our intuitions, can Bayesianism be said to *justify* those intuitions? If the Bayesian has to resort to this *ad hoc* method in order to get the "right' probabilities in the end, one suspects that the reason those observations *ought* to be distinguished has little to do with what the Bayesian *claims* is responsible.

5. Implications for Bayesian Philosophy of Science.

The traditional challenge posed by the Ravens Paradox is to distinguish two sets of data, $X=n$ and $Y=n$, that are logically consistent with a hypothesis. Since the Bayesian finds *a* difference between them that affects the posterior distribution, it may *appear* that this indeed is a success. But it is crucial to see that this difference depends solely on the fact that the likelihood functions, $L_X(\theta)$ and $L_Y(\theta)$, are not proportional. Since $X=n$ and $Y=n$ are

---

[19] This is not to say (A) *must* hold. What we need is an *independent* reason to think (A) fails just when we think $X=n$ and $Z=n$ *ought* to be distinguished.

different sets of data — apples and oranges, as it were — it is not particularly surprising to find such a difference.

No such difference exists when we compare $X=Z=5$, since $L_X(\theta)$ and $L_Z(\theta)$ are proportional. One may be tempted to conclude that because the Bayesian can find no relevant difference between the two, such a difference must *not* exist. Yet this amounts simply to *dismissing* the intuition that there is a relevant evidential difference between $X=5$ and $Z=5$ when the Bayesian sees no difference. This is a step that at least some Bayesians are unwilling to take, and one that seems all the more unmotivated if the alternative account sketched in section 3 is plausible.

The variation on the Ravens Paradox presented above might be taken as just another 'research problem' for Bayesians. But I think it highlights a basic reason to be skeptical of Bayesianism as a framework for understanding how science works. One thing we want in an account of scientific reasoning is an explanation for why, given assumptions that are plausible in a scientific context, certain kinds of methodological judgments are justified. As suggested by the attention the Ravens Paradox has drawn, a particularly important kind of judgment deals with the relationship between how we acquire data and their evidential import. For Bayesians, however, differences in how data sets are generated are either *irrelevant* to what they tell us about hypotheses, or are made relevant only by making what seem to be *ad hoc* assumptions. Given that, it seems that Bayesianism faces a real problem in accounting for an importance kind of methodological judgment in science.[20]

---

[20] Using *non-subjective* prior distributions allow Bayesians to make methods of data-generation relevant to inferences. Such an approach has important virtues, particularly in scientific contexts (Bernardo 1979). However, serious foundational questions surround the use non-subjective priors to represent an agent's *beliefs*, which is typically how Bayesian philosophers of science treat priors (Seidenfeld 1979, Berger and Wolpert 1988).

6. References

Berger, J. and R. Wolpert. 1988. *The Likelihood Principle*, 2nd edition. Hayward, CA: Institute of Mathematical Statistics.

Bernardo, José M. 1979. "Reference Posterior Distributions for Bayesian Inference (with discussion)." *Journal of the Royal Statistical Society, Series B* **41**, 113-147.

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

Edwards, A.W.F. 1992. *Likelihood (expanded edition)*. Baltimore and London: The Johns Hopkins University Press.

Fitelson, B. 1999. "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity." *Philosophy of Science* **66** (Proceedings), pp. S362-S378.

Forster, M. 1995: "Non-Bayesian Foundations for Statistical Estimation, Prediction, and the Ravens Example." *Erkenntnis* **40**: 357-376.

Forster, M. and E. Sober. 1994. "How To Tell When Simpler, More Unified, or Less *Ad hoc* Theories Will Provide More Accurate Predictions." *The British Journal for the Philosophy of Science* **45**: 1-35.

Good, I. J. 1967. "The White Shoe is a Red Herring." *The British Journal for the Philosophy of Science* **17**: 322.

Horwich, P. 1982. *Probability and Evidence*. Cambridge: Cambridge University Press.

—— 1993. "Wittgensteinian Bayesianism." In *Midwest Studies in Philosophy: Philosophy of Science*, vol. XVIII, French, Uehling, Jr., and Wettstein (eds.), Notre Dame, IN: Notre Dame University Press.

Hosiasson-Lindenbaum, J. 1940. "On Confirmation." *Journal of Symbolic Logic* **5**: 133-148.

Howson, C. and P. Urbach. 1992. *Scientific Reasoning: The Bayesian Approach*. 2nd edition. Chicago and La Salle, IL: Open Court.

Mayo, D. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Royall, R. 1997. *Statistical Evidence: The Likelihood Paradigm*. New York: Chapman & Hall.

Seidenfeld, Teddy. 1979. "Why I am Not an Objective Bayesian." *Theory and Decision* **11**, 413-440.