Infinitely Many Resolutions of Hempel's Paradox*

Kevin B. Korb
Dept. of Computer Science
Monash University
Clayton, Victoria 3168
Australia
korb@bruce.cs.monash.edu.au

November 29, 1993

What man is so mad as wil say the swan is black: or that the raven is in couloure white, when the matter is otherwise to be judged by commonsense?

–Cardanus ca. 1576

Abstract. What sorts of observations could confirm the universal hypothesis that all ravens are black? Carl Hempel proposed a number of simple and plausible principles which had the odd ("paradoxical") result that not only do observations of black ravens confirm that hypothesis, but so too do observations of yellow suns, green seas and white shoes. Hempel's response to his own paradox was to call it a psychological illusion—i.e., white shoes do indeed confirm that all ravens are black. Karl Popper on the other hand needed no response: he claimed that no observation can confirm any general statement—there is no such thing as confirmation theory. Instead, we should be looking for severe tests of our theories, strong attempts to falsify them. Bayesian philosophers have (in a loose sense) followed the Popperian analysis of Hempel's paradox (while retaining confirmation theory): they have usually judged that observing a white shoe in a shoe store does not qualify as a *severe* test of the hypothesis and so, while providing Bayesian confirmation, does so to only a *minute* degree. This rationalizes our common intuition of non-confirmation.

All of these responses to the paradox are demonstrably wrong—granting an ordinary Bayesian measure of confirmation. A proper Bayesian analysis reveals that observations of white shoes may provide the raven hypothesis *any degree of confirmation whatsoever.*

Keywords. Confirmation, confirmation paradoxes, Bayesian reasoning, induction, inductive inference, protocol.

*I wish to acknowledge the helpful comments of David Dowe, Brian Ellis, Chris Wallace, John Bigelow, Neil Thomason, Noretta Koertge and participants of the 1993 Australian Association for the History, Philosophy and Social Studies of Science Conference at LaTrobe University.

1 Introduction

There has been a great deal of recent interest in, and controversy about, Bayesian approaches to understanding confirmation and statistical inference. What distinguishes Bayesian from other views is that it takes Thomas Bayes' theorem (explained below) for calculating conditional probabilities as a rule for conditioning the probability of hypotheses upon new evidence, and regards this as a key to any normative account of hypothesis evaluation. Bayesian thought has made substantial progress within the philosophy of science (cf. Howson and Urbach 1989 for an account of much of that) and within artificial intelligence as a method for reasoning under uncertainty (cf. Neapolitan 1990 and Pearl 1988).

Nevertheless, there remain puzzles and anomalies aplenty to occupy Bayesian theorists. One such puzzle is variously known as Hempel's paradox, the raven paradox, and "the" paradox of confirmation. Most Bayesians (including me, below) will claim that the paradox is resolved—*easily*—using Bayesian principles. The remaining difficulty is that these resolutions are non-identical. I intend to finally eliminate disagreement by embracing all possible resolutions of the paradox.

My strategy here is to *assume* that some variety of Bayesianism is correct (I am, after all, some variety of Bayesian!). This is not question-begging, for it is important for advocates and opponents alike to find out just what Bayesian principles commit one to regarding confirmation theory. In accord with this strategy I will blithely assume, for example, that the prior probabilities of hypotheses, needed to operate Bayes' theorem, are available.

2 Bayesian Preliminaries

Bayes' theorem is a non-controversial theorem of the probability calculus:

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}.$$

This asserts that the probability of hypothesis h conditional upon the evidence e is equal to the *likelihood* ($P(e|h)$) times the prior probability of h divided by the probability of the evidence. What is controversial is the further equation of this quantity with the *posterior* probability of h—asserting that the proper probability to adopt is just the prior conditional probability, on the assumption that it is conditioned upon just the evidence we have received. This latter equation is known as the Bayesian rule of conditionalization. Again, I shall not be defending conditionalization here, but assuming it. Following such an assumption we immediately find a proper means of measuring the degree to which some evidence e supports (or disconfirms) the hypothesis h, namely the extent to which the posterior probability of h on e exceeds (or falls short of) its prior probability:

$$S(h|e) =_{df} P(h|e) - P(h).$$

This measure of support is the ordinary one in Bayesian literature. A second measure of support, the ratio of likelihoods e given h over e given not-h, is equally defensible:

$$\lambda(e|h) =_{df} \frac{P(e|h)}{P(e|\neg h)}.$$

It is a simple theorem that the likelihood ratio is greater than one if and only if $S(h|e)$ is greater than zero. $\lambda(e|h)$ can be understood as a degree of support most directly by observing its role in the odds-likelihood version of Bayes' theorem:

$$O(h|e) = \lambda O(h).$$

This simply asserts that the conditional odds on h given e should equal the prior odds adjusted by the likelihood ratio. Since odds and probabilities are interconvertible ($O(h) =_{df} P(h)/P(\neg h)$), support defined in terms of changes in normative odds measures changes in normative probabilities quite as well as $S(h|e)$. λ does have one significant advantage over $S(h|e)$ however: it is simpler to calculate. Indeed, since a likelihood is just the probability of evidence given a hypothesis, and since hypotheses often describe how a causal system functions given some initial condition, finding the probability of the evidence assuming h is often a straightforward computation.[1] What a likelihood ratio reports is the normative *impact* on the posterior probability of the evidence, rather than the posterior probability itself (i.e., the other necessary ingredient for finding the posterior is the *prior* probability of h). However, confirmation theory is primarily concerned with accounting just for rational changes of belief, and so λ is an appropriate measure for dealing with many of its concerns.[2]

3 Hempel's Paradox

To motivate Hempel's paradox historically we need to take a quick look at early logical positivist views on confirmation. The positivists in the early century were deeply impressed by the turn of the century revolutionary developments in the sciences, especially relativity theory (cf. Coffa, 1991). They wanted to account for what was good in science, initially demarcating sense from nonsense (i.e., science from non-science) by insisting that all meaningful sentences be verifiable through direct sensory experience (cf. Hempel, 1965). A little reflection produced the embarrassment that the general hypotheses of the sciences that positivists admired would not be accommodated by this definition. Jean Nicod (1923) proposed therefore a qualitative *positive instance* criterion of confirmation to replace verification:

> e *confirms* a general hypothesis h just in case e is a *positive instance* of h.
>
> If h is $\forall x(Rx \supset Bx)$ (or more simply, $R \rightarrow B$), e is a *positive instance* of h iff for some object a, $\vdash e \equiv Ra \wedge Ba$.[3]

It is surely plausible that positive instances confirm general hypotheses, for this seems to be necessary if *anything* is to confirm general hypotheses.

Carl Hempel (1945) took up Nicod's criterion, with various refinements, and it has remained popular within the philosophy of science ever since (cf. Glymour's bootstrapping criterion of confirmation, 1980). For our purposes, the most important refinement was Hempel's addition of the *equivalence condition*:

> If e confirms h and if h and h' are logically equivalent, then e confirms h'.

This too is a plausible principle: since logically equivalent hypotheses are interchangeable in all but opaque contexts (e.g., I might believe h but not h' because I fail to *realize* that they are logically equivalent), it seems imperative that any evidence support them equally. Hempel's paradox is now immediate. Consider the raven hypothesis: that all ravens are black, or $R \rightarrow B$. This is logically equivalent to its contrapostive $\neg B \rightarrow \neg R$. By Nicod's criterion, any positive instance of $\neg B \rightarrow \neg R$, say white shoes, confirms it. By equivalence, white shoes also confirm that ravens are black!

3.1 Hempel's Answer

Hempel, having found the paradox, thought he also had found the answer: namely, that the feeling of paradox is illusory. By his lights it is perfectly OK to confirm the raven hypothesis with white shoes or anything else that conforms to the hypothesis. The contrary *feeling* arises from the notion that the raven hypothesis is *about* ravens—but that is a wrong notion. The hypothesis $\forall x(Rx \supset Bx)$ is logically equivalent to $\forall x(\neg Rx \wedge Bx)$; thus, the raven hypothesis asserts of *every thing* that it is either not a raven or is black. A white shoe observation therefore exhausts *some* of the content of h. Hempel wrote (1945, p. 18), "The impression of a paradoxical situation is not objectively founded; it is a psychological illusion."

Not very many philosophers of any stripe have followed Hempel on this point (even Glymour defects, opting for an ad hoc rule to exclude white shoes; cf. 1980, pp. 157ff). At least one good reason to differ is that Hempel's analysis fails to explain why naturalists don't conduct their observations at the shopping mall.

4 Bayesian Analyses of the Paradox

A Bayesian response that is less than universal, but still common, may be arrived at via the following line of thought. Bayesians are obliged to assert that Karl Popper was wrong about most things methodological, including of course his denial of the *possibility* of confirmation and the flip side thereof, an overemphasis on falsification (Popper, 1959). But Popper was right about this: severe tests are preferable to bland, common-place predictions.[4] Most hypotheses predict the ordinary, so ordinary predictions can hardly help us distinguish between fruitful and fruitless hypotheses. What we want are predictions that are a priori *surprising*—such as that light 'bends' or that some diseases are inheritable. A severe test then is one in which the outcome e predicted by h would be a surprise were h false. Bayesianism supplies a very natural interpretation of severity.

A test of h with outcome space $\{e, \neg e\}$ is severe just in case:

i) e is predicted by h: $P(e|h) = \text{high}$

ii) e would otherwise be surprising: $P(e|\neg h) = \text{low}$

Jointly, these two conditions imply that

$$\lambda = \frac{P(e|h)}{P(e|\neg h)} = \text{very high.}$$

Severe tests therefore set the stage for outcomes that are highly confirmatory.

What is wrong with shopping mall excursions is just that hunting for, and finding, white shoes does not constitute a severe test for the raven hypothesis. In short: white shoe observations may confirm the raven hypothesis, but only to an *insignificant* degree. This claim saves our intuition that the degree of confirmation is effectively nil and so saves our methodological avoidance of shopping malls. Note too that this response is unavailable to Hempel (and Glymour): Hempel's program was to provide a *qualitative* account of confirmation, and so reference to insignificant degrees of confirmation is unsupportable.

So far so good. But *why* are white shoe observations of low confirmatory power? Asserting this is one thing, justifying it another. The 'standard' Bayesian justification is succinctly expressed by Howson and Urbach (1989). They claim that if we assume a discipline of random (and uniform) selection from among all objects in the universe (we shall have to also assume the universe is finite then), we can ascribe the varying degrees of confirmation afforded by different observations entirely to the varying class sizes of the different relevant types of objects (p. 90):

> ... since non-black, non-ravens form such a numerous class compared with black ravens, it is almost (but not absolutely) certain that a random object about which we know nothing will turn out to be neither black nor a raven, but relatively unlikely that it will be a black raven. Hence, for a Bayesian, both kinds of object confirm 'All ravens are black', but non-black, non-ravens do so only minutely.

This is dead wrong. It is a misapplication of the Bayesian analysis of severity. Even supposing we can employ a discipline of universal random selection, λ is *not* a function *only* of the relative proportions of class sizes relevant to h — λ is also dependent upon the hypothesis space and the experimental protocol used to test h.

4.1 The Correct Bayesian Analysis

The correct Bayesian analysis, whatever it may be, must conform to the verdict provided by λ (equivalently, $S(h|e)$). So far, we have not been given enough information in the raven story to calculate any likelihood; so in some sense the correct analysis must be to announce that we have no idea what the confirmatory impact of white shoe observations might be. However, let us consider what the story lacks for the calculation of $P(e|h)$ and $P(e|\neg h)$. We've already been given a protocol for conducting the observation: uniform random selection among all objects. Taking e to mean that a white shoe is observed, we need also to know then what proportion of all objects are white shoes on the assumption that ravens are black.[5] To make things definite, let us assume (for the moment) that the raven hypothesis has no effect on the supposed number of ravens and also on the supposed number of total objects; indeed, let us assume that the space of possible hypotheses is exactly as given in Figure 1.

In this case, we do have enough information to calculate the relevant likelihoods (where $|U|$ is the number of objects in the universe):

$$P(e|h) = 10^9/|U|$$

$$P(e|\neg h) = 10^9/|U|$$

	R	S
B	10	10^6
W	0	10^9

$h: R \rightarrow B$

	R	S
B	5	10^6
W	5	10^9

$\neg h : \neg[R \rightarrow B]$

Figure 1

So, $\lambda = 1$. White shoe observations do not have a minute impact on the raven hypothesis; *they have no impact whatsoever.*

4.2 Another Bayesian Analysis

Paul Horwich (1982) has provided yet another Bayesian analysis of the raven hypothesis—one that again accounts for the observational protocol. Rather than assume that we randomly select from among *all* objects, Horwich asks what happens when we randomly select from one or another relevant subset of objects. In particular, he asks whether the observation of a black raven might lead to a different evaluative outcome depending upon whether we are selecting (randomly) from among all ravens or from among all black things. He answers, yes! Why? Because R^*B subjects h to the risk of falsification,[6] whereas RB^* does not.[7]

Howson and Urbach pronounce the argument 'specious' (1989, p. 91): "The only difference between R^*B and RB^* is in the point at which one learns that the hypothesis has not been refuted."

Howson and Urbach are wrong; Horwich is right for the wrong reason. Howson and Urbach are being mesmerized by the propositional content of the observation report: it is the same in both cases, therefore presumably its impact on the hypothesis must be the same in both cases. This presumption is in error because *what is important for confirmation*—viz. the likelihoods—is not fully determined by the propositional content of e and h. R^*B and RB^* select the *same* observational statement e, but generate different likelihood ratios. Taking as our example Figure 1 again we find:

$$\lambda(R^*B|h) = \frac{P(R^*B|h)}{P(R^*B|\neg h)} = \frac{1}{.5} = 2$$

Whereas,

$$\lambda(RB^*|h) = \frac{P(RB^*|h)}{P(RB^*|\neg h)}$$

$$= \frac{10/(10+10^6)}{5/(5+10^6)}$$

$$= 2 \times \frac{10^6+5}{10^6+10}$$

$$= 2 - \epsilon$$

Clearly, $\lambda(R^*B|h) > \lambda(RB^*|h)$. But, contra Horwich, this is not due to superior 'falsifiability'. In fact, the B^* observations *can* end up refuting h: consider what would happen if we *exhausted* B^* without finding a raven (when we know a priori that there are ravens in the world).[8]

4.3 Rationalizing the Standard Bayesian Response

None of the above removes the intuition that there is surely *something* right about the original Bayesian point—that white shoe observations speak to the raven hypothesis, but in a vanishingly small voice. It is proper to ask of any candidate confirmation theory whether it can explain—or explain away—our intuitions on such matters. Can Bayesian confirmation theory actually account for this intuition? Yes. All we need is to substitute a new protocol to rationalize this idea:

$$\lambda(W^*S|h) = \frac{P(W^*S|h)}{P(W^*S|\neg h)}$$

$$= \frac{10^9/10^9}{10^9/(10^9+5)}$$

$$= \frac{10^9+5}{10^9}$$

$$= 1 + \epsilon$$

It is not Howson and Urbach's *intuitions* that are at issue here, rather it is their explanation of those intuitions—*which ignores protocol*—that is objectionable.[9]

5 Remarks

We have seen that the observational or experimental protocols whereby we conduct our tests are crucial to understanding the normative impact of evidence on our assessment of hypotheses. This is really little more than common sense. Commonsensically, for example, we would hardly take a known prankster's production of a counterfeit bill as seriously as a randomly discovered counterfeit bill in support of the hypothesis that the number of counterfeits in general circulation is large. Yet both observations may well be reported using the same language. Refusing to acknowledge the varying evidential impact here is indeed "propositional mesmerization."

Now many Bayesians will respond by pointing out that when we observe the prankster produce the bill we are observing much more than the bill alone, and that this information is properly recorded in the statement of the evidence. This may be so. In that case we will obtain a much richer and more complex outcome space, one that records what ordinarily would be thought of as features of experimental procedure rather than outcomes. It makes no difference to λ how those features are taken into account, and so it makes no difference to Bayesian confirmation theory. If others can make explicit in the observation language all that is needed to calculate likelihoods (beyond the hypotheses), then good for them.

What this response should not be allowed to do, however, is provide an excuse for pretending there is no difficulty here to think about. The need for the *explicit* consideration of protocol has not been noticed only recently; for example, Jerzy Neyman (1950) discusses it in the context of selecting a proper reference class. And recent interest in the subject has been stimulated by Shafer (1985). Nevertheless, Bayesianism has an odd history of simply ignoring protocol or of discounting its impact on confirmation. Witness the mishandling of Hempel's paradox.

Another clear case is Bayesian confusion about 'stopping rules': it is standard Bayesian dogma that the rule used to stop the gathering of a sample is irrelevant to statistical inference; all that counts is the sample actually gathered (cf. Howson and Urbach, 1989; pp. 169-171). This is just a special case of the sweeping denial that protocol is relevant. This denial does not cohere with the acceptance of the likelihood ratio as a measure of degree of confirmation.[10]

Another point that needs emphasis, and is often overlooked by Bayesians, is that the hypothesis space is crucial. Without hypotheses alternative to the one under test it is obvious that we will have no likelihood ratio, for we will have no likelihood $P(e|\neg h)$. Therefore, some specific space of hypotheses must be assumed. My own example hypothesis space for the raven hypothesis is very specific indeed: not only do I assume in case h is false that there are non-black ravens, I assume that there are exactly 5 of them! Indeed, it would not be unreasonable to claim that I've simply *rigged* my results in my choice of hypothesis space.[11]

My point is: not only have I done this, but I *can* do this. That is, likelihood ratios are indeed a partial function of the hypothesis space, and therefore sweeping claims about the degree of support that the raven hypothesis properly receives from white shoe observations—*independent of any consideration of protocol and hypothesis space*—are foolishness. Bayesians, when discussing Hempel's paradox, have done so under the guise of presenting a *general* theory of confirmation; if so, such theories must apply to *any* hypothesis space.[12]

6 What Went Wrong with Hempel?

A remaining bit to be tidied up requires a Bayesian analysis of what is wrong with Hempel's understanding of the paradox. Bayesians are obligated not to complain about Hempel's equivalence condition; for if two hypotheses are logically equivalent, then probabilistic consistency requires that they be assigned identical probabilities, etc. What is wrong is the idea that positive instances *necessarily* are confirming instances. There's no denying that they *commonly* are confirming instances (if positive instances seldom confirmed, then what would?); but it is a long way from

	R	S
B	1	1
W	0	2

$h : R \rightarrow B$

	R	S
B	1	1
W	1	1

$\neg h : \neg[R \rightarrow B]$

Figure 2

commonality to universality.

6.1 Counterexample 1: Hexed Salt (Salmon, 1971)

It is typical that the probability functions we are concerned with already embed (via conditionalization, for example) a good deal of background knowledge. Here I shall represent the background knowledge K explicitly as a subscript. Suppose that in our background knowledge we are aware that all salt dissolves. Then we shall hardly expect hexed salt to do otherwise, and we will not take a positive instance of hexed salt dissolving as confirmatory for the hypothesis that hexed salt dissolves. Bayesianism trivially accommodates this point: since $P_K(e) = 1$, $\lambda_K(e|h) = 1$; i.e., e is non-confirmatory because, although $P_K(e|h) = 1$, $P_K(e|\neg h) = 1$ also.

6.2 Counterexample 2: Killer Bees (Swineburne, 1971)

Let h assert that killer bees cannot live north of the 38th parallel north. Let e assert that killer bees have been observed at 37.99°N. This is a positive instance of the hypothesis that *disconfirms* h. (Cf. also Good, 1967.)

6.3 Counterexample 3: Black Ravens Are White Shoes

A black raven observation may *fail* to confirm $R \rightarrow B$! This must be, on my analysis, since such observations after all stand to $\neg B \rightarrow \neg R$ in just the same relationship as does the white shoe observation to the raven hypothesis.

Consider then random selection from U and Figure 2. In this hypothesis space, observing a randomly selected black raven fails to support the raven hypothesis. If this seems counterintuitive, think of h and $\neg h$ as two urns with white and black balls. Retrieving a black ball marked R fails to help us in deciding whether the sample has come from the urn h or the urn $\neg h$. If your incredulity applies to the idea of extending this point to hypotheses about *ravens*, then I think you must be right: this kind of hypothesis space wouldn't seem to fit our world very well; in any case, random selection from the *universe* is not a possible protocol to implement.

7 Conclusion

We should well and truly forget about positive instance confirmation: it is an epiphenomenon of Bayesian confirmation. There is no qualitative theory of confirmation that can adequately approximate what likelihood ratios tell us about confirmation; nor can any qualitative theory lay claim to the success (real, if limited) of Bayesian confirmation theory in accounting for scientific methodology.

Bayesian confirmation theory, on the other hand, cannot even begin without an *explicit* consideration of observational or experimental protocol and hypothesis space. Propositional mesmerization—an overconcern with how experimental outcomes are *recorded*, rather than with what they imply within their experimental and hypothetical context—has led to confused claims about scientific methodology. Whereas such confusions pose little threat to scientific practice, since practitioners will for the most part ignore philosophers of science, there is need for greater clarity when it comes to developing computational models of Bayesian reasoning, whether the models be intended descriptively or normatively. It is a virtue of such modeling that it requires complete explicitness, thereby providing a difficult test of the clarity and coherence of the philosophical basis for one's model of inference. I believe I have shown that a Bayesian model that simply ignores protocol will fail that test.

Notes

1. I must confess straightaway that finding the *other* likelihood $P(e|\neg h)$ is often much more problematic, since $\neg h$ is likely to be much more complicated than a single alternative causal hypothesis. But pursuing this important issue would yield a completely different, if related, paper; so I shall continue to rely upon nonchalance here.

2. Nor am I the first Bayesian to focus on λ rather than $S(h|e)$: I.J. Good has primarily concerned himself with likelihood ratios (in log form; Good, 1950).

3. Nicod also considered any instance a such that $\neg Ra$ to be evidentially irrelevant to h; but I ignore this since Hempel and others abandoned that aspect of Nicod's criterion.

4. A great many philosophical Bayesians are immigrants from Popperian territory (including me).

5. What is protocol? Shafer (1985, p. 261) defines it as the "set of rules that tell, at each step [of the experiment], what can happen next." Perhaps more vaguely, I will take as protocol whatever needs to be known, in addition to the hypothesis space, in order to calculate the likelihood ratios relevant for some test.

6. To simplify expression, we can let the two distinct ways of arriving at the same observation statement be recorded thus: 'R^*B' shall mean that a black raven has been observed when selecting among ravens, while 'RB^*' shall mean we have the same observation when selecting from among all black objects.

7. Horwich is clearly another Popperian turned Bayesian.

8. Some have wondered whether I am being quite fair to Horwich here. My assumptions here are at least not obviously unfair. It is *Horwich's* assumption that B be exhaustible, since otherwise the protocol of random, uniform selection therefrom is impossible. And no participant in these discussions has yet volunteered to doubt that there are indeed ravens. If it requires only the

drawing upon such an evident fact to find fault with Horwich's methodological pronouncements, then surely something is wrong with those pronouncements.

9. Although the numbers of Figure 1, which my rationalization here depends upon, do not make sense interpreted literally, the acceptability of my explanation depends only upon the *proportions* being more or less sensible.

10. In support of my assertion, consider this example. We are interested in testing the hypothesis h, that a certain coin is biased s.t. the probability of heads is 2/3, against h', that the coin is fair. Let the evidence that is reported be e: that in the sample of flips obtained the number of heads is equal to the number of tails. Is protocol, in the form of a stopping rule, irrelevant to how this evidence should be assimilated? Hardly.

Protocol A. Flip the coin twice. In this case,

$$\lambda(e|h) = \frac{P(e|h)}{P(e|\neg h)} = \frac{2!(2/3)(1/3)}{2!(1/2)^2} = \frac{8}{9}$$

Protocol B. Flip the coin until the sample has equal numbers of heads and tails. Random walk theory tells us that $P(e|h') = 1$. Random walk theory further tells us that the probability of walking away from an absorbing (stopping) state is zero if the bias away from that state is zero or negative; otherwise, if the bias is positive and represented by $\frac{p}{q} > 1$, then the probability of absorption is $\left(\frac{q}{p}\right)^a$, where a is the number of steps (with q probability) away from the stopping state (cf. Cox and Miller, 1965; §2.2). Assuming h, after one flip we shall be in state -1 with probability (1/3) or in state $+1$ with probability (2/3). The probability of absorption is one in the first case and (1/2) in the second; therefore, $\lambda(e|h) = P(e|h) = \frac{1}{3} + \frac{2}{3} \times \frac{1}{2} = \frac{2}{3} \neq \frac{8}{9}$.

Although I like protocol B, if anyone is disconcerted by protocols that may never complete, we get a similar result with:

Protocol C. Flip the coin four times. We have,

$$\lambda(e|h) = \frac{P(e|h)}{P(e|\neg h)} = \frac{\binom{4}{2}(2/3)^2(1/3)^2}{\binom{4}{2}(1/2)^4} = \left(\frac{8}{9}\right)^2 \neq \frac{8}{9}$$

11. See Earman (1992) for an interesting, if inconclusive, discussion of the Bayesian practice of ignoring the hypothesis space.

12. And here is where the proof of my title would come in. I leave this as an exercise for the reader, with the obvious hint that to obtain arbitrary likelihood ratios for white shoes be prepared to apply further contortions to hypothesis and protocol spaces.

References

Coffa, J. Alberto (1991) *The Semantic Tradition from Kant to Carnap: To the Vienna Station.* Cambridge University.

Cox, D.R. and H.D. Miller (1965) *The Theory of Stochastic Processes.* London: Methuen.

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

Glymour, C. (1980) *Theory and Evidence.* Princeton University.

Good, I.J. (1950) *Probability and the Weighing of Evidence.* London: Charles Griffin.

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

Hempel, C. (1945) "Studies in the Logic of Confirmation," pp. 3-46 of C. Hempel *Aspects of Scientific Explanation*, New York: Free Press (1965).

Hempel, C. (1965) "Empiricist Criteria of Cognitive Significance: Problems and Changes," pp. 101-119 of C. Hempel *Aspects of Scientific Explanation*, New York: Free Press (1965).

Horwich, P. (1982) *Probability and Evidence.* Cambridge University.

Howson, C. and P. Urbach (1989) *Scientific Reasoning: The Bayesian Approach.* La Salle, Ill.: Open Court.

Neapolitan, R. (1990) *Probabilistic Reasoning in Expert Systems: Theory and Algorithms.* New York: Wiley.

Neyman, J. (1950) *First Course in Probability and Statistics.* New York: Holt.

Nicod, J. (1923) "The Logical Problem of Induction," translated by M. Woods in J. Nicod *Geometry and Induction*, Berkeley: University of California (1970).

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems.* San Mateo: Morgan Kaufmann.

Popper, K.R. (1959) *The Logic of Scientific Discovery.* New York: Basic Books.

Salmon, W.C. (1971) *Statistical Explanation and Statistical Relevance.* University of Pittsburgh.

Shafer, G. (1985) "Conditional Probability," *International Statistical Review 53*: 261-277.

Swineburne, R.G. (1971) "The Paradoxes of Confirmation—A Survey," *American Philosophical Quarterly 8*: 318-329.