Likelihoodism, Bayesianism, and Relational Confirmation

Branden Fitelson*
University of California/Berkeley

01/01/05

Abstract. Likelihoodists and Bayesians seem to have a fundamental disagreement about the proper probabilistic explication of relational (or contrastive) conceptions of evidential support (or confirmation). In this paper, I will survey some recent arguments and results in this area, with an eye toward pinpointing the nexus of the dispute. This will lead, first, to an important shift in the way the debate has been couched, and, second, to an alternative explication of relational support, which is in some sense a "middle way" between Likelihoodism and Bayesianism. In the process, I will propose some new work for an old probability puzzle: the "Monty Hall" problem.

Keywords: Confirmation, Support, Favoring, Likelihood, Bayesian, Monty Hall

1. Introduction: Setting the Stage for the Contemporary Debate

There are various evidential concepts that have received probabilistic explications in the past half-century or so. Carnap (1962) provides an early (but still quite useful) taxonomy, consisting of three main types of confirmation: classificatory (or qualitative), comparative (or relational), and quantitative. Here's an excerpt from (Carnap, 1962).[1]

> ***The comparative concept of confirmation*** is is usually expressed in sentences of the following or similar forms:
>
> > 'H_1 is more strongly confirmed (or supported, substantiated, corroborated, etc.) by E_1 than H_2 is by E_2.'

Carnap's "comparative confirmation" concept comes close to the relational support concept that will be our focus presently. But, there are two important differences. First, we will only be concerned with a special case of Carnap's comparative concept in which there is just one

* Thanks to the participants of the Philosophy, Probability, and Modeling (PPM) Seminar at the University of Konstanz (especially Stephan Hartmann, Franz Huber, Wolfgang Spohn, and Teddy Seidenfeld), for a very fruitful discussion of an early draft of this paper in July, 2004. Since then, discussions and correspondences with Prasanta Bandyopadhyay, Luc Bovens, Alan Hájek, Jim Hawthorne, Jim Joyce, Jon Kvanvig (and other participants of his "Certain Doubts" blog, which had a thread on a previous draft of this paper), Patrick Maher, Sherri Roush, Richard Royall, Elliott Sober, Dan Steel, and an anonymous referee of *Synthese* has been very valuable.

[1] Here, I have altered Carnap's notation, to make it more consistent with ours.

evidential proposition E and two alternative hypotheses H_1 and H_2. Second, and more importantly, some contestants in the contemporary debate about relational support take issue with the implicit presupposition Carnap makes in his classification: that *relational* claims like "evidence E favors hypothesis H_1 over hypothesis H_2" can be reduced to a comparison of *non*-relational confirmational quantities (*i.e.*, the degree to which E non-relationally confirms H_1 and the degree to which E non-relationally confirms H_2). Likelihoodists such as Elliott Sober (1994) and Richard Royall (1997) reject this reductive presupposition. They prefer a non-reductive, Likelihood-based account of a primitive, three-place favoring relation. Contemporary Bayesians, on the other hand, tend to adhere to a modern form of Carnapian reductionism. Bayesians [*e.g.*, Peter Milne (1996)] typically think of relational confirmation as a derived concept, which is defined in terms of their primitive, non-relational confirmation concept. In a nutshell, this is the main locus of disagreement between the Bayesian and Likelihoodist approaches we will discuss. While it is tempting (and Carnapian) to paraphrase "E favors H_1 over H_2" as (something like) "E supports H_1 more strongly than E supports H_2", we must not assume at the outset that relational support (favoring) claims can or should be reduced to a comparison of non-relational confirmation claims. Whether such a reduction is possible (or, preferable) is one of the questions this paper will address.

In the next section, I will describe the Likelihoodist account of favoring. In section three, I will discuss Bayesian accounts of favoring. In section four, I will critically examine two recent Likelihoodist arguments against (reductive) Bayesian accounts of favoring. Here, I will argue that Likelihoodists have not met their argumentative burden, and that Bayesian accounts of favoring capture an important aspect of favoring that even Likelihoodism cannot responsibly ignore. In the end, however, I will not try to defend my favorite Bayesian account of favoring. Instead, I will describe both Bayesian and non-Bayesian alternatives to Likelihoodism, including an alternative non-Bayesian approach which is somewhere "in between" Likelihoodism and traditional Bayesianism. My main goal in this paper is simply to clarify what's at issue between Likelihoodists and non-Likelihoodists, and to explain what I think needs to be done, from a Likelihoodist point of view, to argue in favor of Likelihoodism – in contrast to its most promising alternatives. Along the way, I will give a novel and illustrative confirmational analysis of an old chestnut: The "Monty Hall" Problem. It turns out that the Monty Hall problem is very well-suited to testing a wide variety of alternative theories of favoring against each other.

2. Likelihoodism, Favoring, and the "Law" of Likelihood

As Sober (1994) explains, Likelihoodists are *contrastive empiricists*. They believe that evidential support is inherently *relational* in a way that Carnap's comparative confirmation concept is not. Here is a representative excerpt from (Sober, 1994, his emphasis, my brackets):

> ... theory testing is a contrastive activity. If you want to test a theory T, you must specify a range of alternatives – you must say what you want to test T *against*. There is a trivial reading of this thesis that I do not intend. To find out if T is plausible is simply to find out if T is more plausible than *not*-T. I have something more in mind: there are various contrasting alternatives that might be considered. If T is to be tested against T', one set of observations [E] may be needed, but if T is to be tested against T'' a different set of observations [E'] may be needed. By varying the contrasting alternatives, we formulate genuinely different testing problems.

Likelihoodists propose a very simple and elegant probabilistic account of their contrastive favoring relation, embodied in the so-called "Law of Likelihood" [this terminology is used by various Likelihoodists, including Edwards (1992), and Royall (1997)]:

(LL) Evidence E favors hypothesis H_1 over hypothesis H_2 *if and only if* ($\iff$) H_1 confers greater probability on E than H_2 does.

This first, informal statement of (LL) is rather rough. To make it more precise, we need to clarify what is meant by the locution "H_1 confers a greater probability on E than H_2 does". Presumably, the (LL) is called the Law of *Likelihood* because it takes the relationship between the *likelihoods* of the hypotheses [$\Pr(E \mid H_1)$ and $\Pr(E \mid H_2)$] as constitutive of the favoring relation.[2] And, presumably, the (LL) is called the *Law* of Likelihood because Likelihoodists think it holds *generally* and not just in certain special cases. On its face, the (LL) might seem somewhat reasonable. After all, shouldn't it be differences in what alternative hypotheses "say about the evidence" that determines which hypothesis is favored by said evidence? And, doesn't the likelihood of H [$\Pr(E \mid H)$]

[2] Likelihoodists are not always consistent on this point. Sometimes, as in (Sober, 1994), Likelihoodists talk about likelihoods as *conditional* probabilities (relative to a background probability model $\mathcal{M}$) of E on H_1 *vs* H_2. But, sometimes, as in some passages of (Royall, 1997), they seem to think of them as *unconditional* probabilities (of E) *entailed by* H_1 *vs* H_2. It is only on the former reading that Likelihoodism and Bayesianism can to come into conflict. On the latter reading, Likelihoodism is basically a *deductivist* theory in which alternative hypotheses *entail* different unconditional probabilities for the evidence. This is incommensurable with Bayesianism *on its face*, since Bayesianism appeals only to *inductive* relations *within* a *single* probability model. It is for this reason that I adopt the former reading of the (LL). If one insists on the latter reading (sometimes apparently adopted by Royall and others), then the conceptual gulf between Bayesians and Likelihoodists becomes *even wider*, which only *bolsters* the dilemma for Likelihoodism that I outline below.

fully capture what H "says about" E? These are the sorts of questions that Likelihoodists think are salient for seeking a proper explication of the favoring relation.[3] *Prima facie*, this does not seem unreasonable.

While the (LL) is simple, elegant, and not lacking in *prima facie* intuitive appeal, it is certainly not beyond reproach. Various alleged counterexamples to (LL) have been floated. I will begin with an example that Elliott Sober (2005) now seems to think presents a problem for the (⇒) direction of the (LL).[4] Sober asks us to

> ... suppose one observes (E) that an ace has just been drawn from a deck of cards, and, the hypotheses under evaluation are H_1 = the card is the ace of hearts, and H_2 = the card is the ace of spades or the ace of clubs. From the information, we know that $\Pr(H_1 \mid E) = 1/4$ and $\Pr(H_2 \mid E) = 1/2$, while $\Pr(E \mid H_1) = \Pr(E \mid H_2) = 1$. It would be odd to maintain that the observation does not favor H_2 over H_1; in this case, the favoring relation is mediated by the probabilities of hypotheses, not their likelihoods.

It is interesting that Sober locates the source of the anti-(LL) aspect of this example in the relationship between the *posteriors* of H_1 and H_2 [*i.e.*, in the fact that $\Pr(H_2 \mid E) > \Pr(H_1 \mid E)$ which runs counter to the likelihood ordering $\Pr(E \mid H_2) \ngtr \Pr(E \mid H_1)$]. There is another way to explain why intuitions tend to run counter to the (LL) in such examples. Note that, in such examples, we also have $\Pr(E \mid {\sim}H_1) > \Pr(E \mid {\sim}H_2)$. This is another salient relational property of E, H_1, and H_2. And, it's one that involves *catch-all likelihoods* (or *catch-alls*, for short), *not* posteriors. I will argue, below, that this other, catch-all likelihood inequality undergirds the correct diagnosis of what's going on in this example (and related examples). But, first, I want to argue that Leeds' example is not conclusive. It seems to me that a Likelihoodist could plausibly reply to Leeds by pointing out that – since the likelihoods of the two hypotheses in this example are equal – any probabilistic difference

[3] Likelihoodists typically assume not only the (LL), but also the *Likelihood Principle* (LP), which says (roughly) that $\Pr(E \mid H)$ fully captures "what H says about E." See (Edwards, 1992) and (Royall, 1997) for traditional Likelihoodist discussions of the (LP). I will not be discussing the (LP) presently. For an incisive discussion of the (LP) from a Bayesian confirmation-theoretic perspective, see (Steel, 2003). Steel's discussion serves to bolster my present arguments by showing that all of the Bayesian approaches to confirmation I will discuss *respect* the (LP). On the other hand, as we'll see below, many of these Bayesian approaches *violate* the (LL). As such, the alleged connection between the (LL) and the (LP) is much weaker than many Likelihoodist discussions might lead one to believe. If Steel is right (I think he is), then there is no need for me to address "(LP) therefore (LL)" arguments, since they are fallacious from our present Bayesian confirmtation-theoretic point of view.

[4] This kind of example was originally posed (as far as I know) as a counterexample to the (LL) by Steve Leeds (2000). Such examples are highlighted in Sober's (2005) recent discussion of the (LL). I will refer to this as "Leeds' Example".

between H_1 and H_2 in this example must *merely* be because of a difference between the *priors* of H_1 and H_2. In this sense, it must be the difference between the priors of H_1 and H_2 which explains *both* $\Pr(H_2 \mid E) > \Pr(H_1 \mid E)$, *and* $\Pr(E \mid \sim H_1) > \Pr(E \mid \sim H_2)$ in this example. And, plausibly, *non*-relational properties of H_1 and H_2 (*e.g.*, their *priors*) can not be be determinative of the *contrastive* evidential fact that E favors H_2 over H_1. So, it seems to me that Likelihoodists needn't be swayed by such examples. Nonetheless, Leeds' example and Sober's discussion of it are suggestive. They raise the possibility of understanding "E favors H_1 over H_2" as a claim about the posteriors of H_1 and H_2 [*i.e.*, as $\Pr(H_1 \mid E) > \Pr(H_2 \mid E)$]. I will return to this suggestion later, and I will show that it is a woefully inadequate explication of the favoring relation. But, first, I need to discuss another, more compelling, class of counterexamples to the (LL).

There are more compelling counterexamples to (LL), which are clearly grounded in *relational*, *logical* asymmetries in the way E bears on H_1 *vs* H_2, and *not* in *non*-relational properties (*i.e.*, the priors) of H_1 and H_2. My favorite example is as follows. Again, we're going to draw a single card from a standard (well-shuffled) deck. This time, E = the card is a spade, H_1 = the card is the ace of spades, and H_2 = the card is black. In this example (assuming the standard probability model of card draws), $\Pr(E \mid H_1) = 1 > \Pr(E \mid H_2) = \frac{1}{2}$, but it seems absurd to claim that E favors H_1 over H_2, as is implied by the (LL).[5] After all, *E guarantees the truth of H_2*, but E provides only non-conclusive evidence for the truth of H_1. This suggests the following principle:

(*) If E provides conclusive evidence for H_1, but non-conclusive evidence for H_2 (where it is assumed that E, H_1, and H_2 are all contingent claims), then E favors H_1 over H_2.

Principle (*) is highly intuitive, but one might complain that it only makes sense for a theory of favoring based on posterior probabilities, not likelihoods. As we will soon see, this is untrue — it misses the mark for the same reason that Sober's diagnosis of Leeds' example misses the mark. A more accurate characterization of favoring theories that satisfy (*) is that they are are dependent not only on likelihoods, but also on *catch-alls*. A careful examination of the most promising Bayesian alternatives to Likelihoodism (undertaken below) reveals that they exhibit just this kind of dependence on catch-alls. As far as I know, this sort of example [or the principle (*) that it suggests] has never been directly addressed by Likelihoodists. But, as will become clearer below,

[5] Note: this example is *simultaneously* a counterexample to *both* the (⇒) *and* the (⇐) directions of the (LL). This is another sense in which this example is more compelling than Leeds', which is only relevant to the (⇒) direction of the (LL).

this is just the sort of example that most perspicuously and intuitively separates Likelihoodism from its toughest Bayesian competitors.

So, how do Likelihoodists respond to examples (like Leeds' or mine) that seem to run counter to the (LL)? Some, like Royall, seem to simply ignore them (Royall does not discuss examples of either kind). Presumably, such Likelihoodists feel that there is no need to respond to such examples. Others, like Sober, do feel the pull of examples like these. Indeed, because of examples like Leeds', Sober no longer seems to think of the (LL) as a "law", but more of a "*ceteris paribus* law". He now says (Sober, 2005), concerning examples like Leeds' (and, presumably, he would now say something similar about my example):

> The Law of Likelihood should be restricted to cases in which the probabilities of hypotheses are not under consideration (perhaps because they are not known or are not even "well-defined") and one is limited to information about the probability of the observations given different hypotheses.

The view now seems to be that the (LL) is (strictly) true only in a restricted class of cases in which we don't know the prior (and, hence, posterior) probabilities of the hypotheses in question, but only their likelihoods. This is a perplexing response, for three reasons. First, this seems just to concede that the (LL) is not a law at all (*i.e.*, that, strictly speaking, the (LL) is *false*, which is all the examples aimed to show in the first place). Second, it doesn't get the *ceteris paribus* condition right. As I will explain shortly, it is not the *priors* that need to be unknown here, but the *catch-alls* $\Pr(E \mid {\sim}H_1)$ and $\Pr(E \mid {\sim}H_2)$ that need to be unknown. For it is the catch-alls, not the priors, that, together with the likelihoods, determine the true favoring relations. Third, this response forces the Likelihoodist into a rather uncomfortable dilemma – by their own epistemological lights (more on this dilemma below). Before getting to the bottom of all of this, we first turn to Bayesian accounts of favoring, and to two recent criticisms of them due to Royall.

3. Bayesian Accounts of Favoring

Contemporary Bayesians [*e.g.*, Peter Milne (Milne, 1996)] offer reductive accounts of favoring, along the lines of Carnap's "comparative confirmation" approach (but with a modern, non-Carnapian underlying conception of non-relational confirmation):

(†) Evidence E favors hypothesis H_1 over hypothesis H_2 *if and only if* ($\Longleftrightarrow$) E confirms H_1 more strongly than E confirms H_2.

This statement of Bayesian reductionism (†) is rather rough (it will be made more precise, below). The basic idea here is that there is some *primitive*, *non*-relational degree to which E confirms each of H_1 and H_2 *individually*, and the favoring relation is *defined* in terms of a comparison between these primitive, confirmational quantities.

For contemporary Bayesians, confirmation is a matter of *probabilistic relevance*. Thus, degree of confirmation is measured using some *relevance measure* $\mathfrak{c}(H, E)$ of the "degree to which E raises the probability of H". Various relevance measures have been proposed and defended in the recent literature [see (Fitelson, 1999) and (Fitelson, 2001b) for surveys]. Here are the three most popular Bayesian relevance measures of non-relational confirmation.[6]

- *Difference*: $d(H, E) \overset{\text{def}}{=} \Pr(H \mid E) - \Pr(H)$
- *Ratio*: $r(H, E) \overset{\text{def}}{=} \dfrac{\Pr(H \mid E)}{\Pr(H)}$
- *Likelihood-Ratio*: $l(H, E) \overset{\text{def}}{=} \dfrac{\Pr(E \mid H)}{\Pr(E \mid \sim H)}$

Plugging a relevance measure $\mathfrak{c}$ of non-relational confirmation into (†) yields a corresponding account of favoring. Thus, the schema:

($\dagger_{\mathfrak{c}}$) Evidence E favors hypothesis H_1 over hypothesis H_2, according to measure $\mathfrak{c}$[7], *if and only if* ($\iff$) $\mathfrak{c}(H_1, E) > \mathfrak{c}(H_2, E)$.

For instance, plugging d into ($\dagger_{\mathfrak{c}}$) yields ($\dagger_d$), *etc.* As it turns out, ($\dagger_d$), ($\dagger_r$), and ($\dagger_l$) disagree radically on the nature of favoring. To be more specific, ($\dagger_r$) has a decidedly Likelihoodist character, while ($\dagger_d$) and ($\dagger_l$) are "less Likelihoodist, and more Bayesian". In fact, ($\dagger_r$) is *so* Likelihoodist that it is *logically equivalent* to the (LL)![8] Other Bayesian precisifications of (†) have still different behavior. I won't attempt a general survey here, but the discussion below will reveal some interesting

[6] Sometimes, logarithms of ratio relevance measures are taken, to ensure that they are positive in cases of confirmation, negative in cases of disconfirmation, and zero in cases of irrelevance. Since logs are monotone increasing functions, this has no effect on the *ordinal structure* (see footnote 7) of the resulting accounts of favoring. So, no loss of generality would result in adding logs to our ratios. But, for present purposes, this would only complicate matters, which is why we have no logs.

[7] Relevance measures are to be identified here by their *ordinal* structure. Two relevance measures $\mathfrak{c}_1$ and $\mathfrak{c}_2$ are *ordinally equivalent* iff, for all E_1, E_2, H_1, and H_2, $\mathfrak{c}_1(H_1, E_1) \geq \mathfrak{c}_1(H_2, E_2)$ iff $\mathfrak{c}_2(H_1, E_1) \geq \mathfrak{c}_2(H_2, E_2)$. If two relevance measures are ordinally equivalent, then, as far as we are concerned, they are identical. So, when we say "according to $\mathfrak{c}$", we really mean "according to any measure ordinally equivalent to $\mathfrak{c}$". Thus, ($\dagger_{\mathfrak{c}}$) really denotes an ordinal equivalence class of theories.

[8] Proof: By Bayes's Theorem, $r(H_1, E) > r(H_2, E)$ iff $\frac{\Pr(E \mid H_1)}{\Pr(E)} > \frac{\Pr(E \mid H_2)}{\Pr(E)}$ iff $\Pr(E \mid H_1) > \Pr(E \mid H_2)$. In fact, Peter Milne (1996) – a Bayesian who *accepts* (LL) – uses this property of r to argue that r is the "one true measure" of non-relational confirmation (thus, seeing a possible *modus tollens* as a *modus ponens*). See footnotes 22 and 25 for more on this crucial underlying Bayesian dispute.

differences between the three most popular Bayesian theories of favoring. While the various Bayesian theories of relational confirmation have many differences, they also have some commonalities. The most interesting commonality for our present purposes has to do with the following *sufficient* (but not necessary) condition for favoring:

(WLL) Evidence E favors hypothesis H_1 over hypothesis H_2 *if* $(\Leftarrow)$ $\Pr(E \mid H_1) > \Pr(E \mid H_2)$ *and* $\Pr(E \mid {\sim}H_1) \leq \Pr(E \mid {\sim}H_2)$.

I borrow the name of this condition from Joyce (2004a) who discusses it under the rubric "The Weak Law of Likelihood". The (WLL) is so-called because it is (obviously) strictly logically weaker than the (LL). Therefore, the (WLL) is not something that Likelihoodists can consistently claim to be *false*. It is obvious, therefore (see footnote 8), that $(\dagger_r)$ entails the (WLL). It is only slightly less obvious that $(\dagger_l)$ entails the (WLL).[9] What is not so obvious is that (WLL) is implied by *all* contemporary Bayesian theories of favoring $(\dagger_\mathfrak{c})$![10] Therefore, the (WLL) captures a crucial common feature of all Bayesian conceptions of relational confirmation. It is important to recognize that the way in which the (WLL) transcends the (LL) is not by its dependence on priors or posteriors, but by its dependence on *catch-alls*. I will return to this aspect of the (WLL) later.

Before continuing with a discussion of Royall's two recent critiques of Bayesian favoring, I must digress, briefly, to consider a more traditional Bayesian approach to confirmation and favoring. Carnap (1962) distinguished two concepts of non-relational confirmation: *confirmation as firmness* and *confirmation as increase in firmness*. The latter is just the contemporary, relevance-based Bayesian conception of confirmation, which takes confirmation to be *probability-raising*. The former is a threshold concept, which explicates "E confirms H" as $\Pr(H \mid E) > r$, for some threshold value r. On this view, the posterior probability itself is taken as the measure of degree of non-relational confirmation. And, the corresponding reductive definition of favoring would then be given by the following:

(‡) E favors H_1 over H_2 *if and only if* $\Pr(H_1 \mid E) > \Pr(H_2 \mid E)$.

It seems that Sober has something like (‡) in mind in his discussion of Leeds' proposed counterexample to (LL). But, (‡) is an inadequate Bayesian theory of favoring, because the underlying notion of confirmation on which it is based ignores probabilistic relevance. Consider any

9 By simple algebra, if $\Pr(E \mid H_1) > \Pr(E \mid H_2)$ *and* $\Pr(E \mid {\sim}H_1) \leq \Pr(E \mid {\sim}H_2)$, then $\frac{\Pr(E \mid H_1)}{\Pr(E \mid {\sim}H_1)} > \frac{\Pr(E \mid H_2)}{\Pr(E \mid {\sim}H_2)}$, which is equivalent to $l(H_1, E) > l(H_2, E)$.

10 *Dozens* of Bayesian relevance measures $\mathfrak{c}$ have been proposed and defended in the literature (Fitelson, 2001b). And, *all* of these are such that $(\dagger_\mathfrak{c})$ entails (WLL). This is mentioned in passing by Joyce (2004a). He and I both omit the proofs.

case in which E raises the probability of H_1, but E lowers the probability of H_2. Intuitively, this is a case in which E indicates that H_1 is true, but E indicates that H_2 is false. In such a case, it seems obvious that E provides better evidence for the truth of H_1 than for the truth of H_2. And, as a result, it seems clear that, in such cases, we should say that E favors H_1 over H_2. Obviously, any contemporary, relevance-based Bayesian ($\dagger_{\mathfrak{c}}$) theory of favoring will have this consequence. And, since (as we saw above) the (LL) is equivalent to ($\dagger_r$), Likelihoodism also has this consequence. Unfortunately, (‡) does not have this consequence. In fact, according to (‡), E can favor H_2 over H_1 in such cases, which is absurd.[11] For this reason, nobody defends (‡) anymore, which shows that Sober's diagnosis of Leeds' example must be off the mark [and so is the analogous reaction to principle (*)].[12] While Sober is right that some may be tempted to conflate (‡) and (†), this is only because it is easy to conflate degree of *confirmation* with degree of *probability* (or degree of *belief*). And, it is not degree of probability (or degree of belief) that is relevant to Bayesian accounts of favoring, but degree of confirmation. Likelihoodists are right to point out that these two concepts must be carefully distinguished (and to chide people for not doing so). But, this is no problem for contemporary Bayesian confirmation theory, which resepcts this distinction.

Next, we turn to two recent critiques of contemporary Bayesian accounts of favoring, due to Richard Royall. Here, the dialectic between Likelihoodists and Bayesians will become clearer, and the informal renditions of the salient principles will be made more precise.

---

[11] Concrete example: a natural number n is to be selected at random from the first 10 natural numbers (*i.e.*, n is a random natural number between 1 and 10, inclusive). Let E: $n \in \{1, 2, 8, 9, 10\}$, H_1: $n \in \{1, 2\}$, and H_2: $n \in \{2, 3, 4, 5, 6, 7, 8, 9\}$. In this case, $\Pr(H_1 \mid E) = \frac{2}{5} > \Pr(H_1) = \frac{1}{5}$, $\Pr(H_2 \mid E) = \frac{3}{5} < \Pr(H_2) = \frac{4}{5}$, but $\Pr(H_2 \mid E) > \Pr(H_1 \mid E)$. So, according to (‡), E favors H_2 over H_1, but according to all other theories of favoring discussed in this paper, E favors H_1 over H_2. Note: $\Pr(E \mid H_1) = 1 > \Pr(E \mid H_2) = \frac{3}{5}$. See (Popper, 1954) for similar examples.

[12] The (WLL) has a "sister principle", which differs from the (WLL) by only a symmetric flipping of inequality signs [it's a kind of "dual" of the (WLL)]:

(WLL′) If $\Pr(E \mid H_1) \geq \Pr(E \mid H_2)$ and $\Pr(E \mid {\sim}H_1) < \Pr(E \mid {\sim}H_2)$, then E favors H_1 over H_2.

All (non-Likelihoodist) Bayesian relevance measures $\mathfrak{c}$ are such that ($\dagger_{\mathfrak{c}}$) entails (WLL′). Thus, all contemporary (non-Likelihoodist) Bayesian theories of favoring say that E favors H_2 over H_1 in Leeds' example. *Pace* Sober, *this* is the source of the anti-(LL) sentiments raised by Leeds' example. Sober's (‡)-based explanation is less diagnostic, since the antecedent of (WLL′) does not entail $\Pr(H_2 \mid E) > \Pr(H_1 \mid E)$. This *explains why* Sober's diagnosis of the Leeds phenomenon [and (*)] is off the mark. It's not sensitivity to *posteriors* that distinguishes Bayesian from Likelihoodist theories of favoring. Rather, it's sensitivity to *catch-alls*, as in (WLL) and (WLL′).

4. Richard Royall's Two Critiques of Bayesian Favoring

Richard Royall (Royall, 1997, ch. 1) offers several critiques of Bayesian theories of favoring. I will discuss only two of these (but, these two critiques are representative of the kinds of criticisms Likelihoodists have made, historically, of Bayesian accounts of relational support). The first critique is aimed specifically against ($\dagger_r$).[13] Royall considers an example involving diagnostic testing. The evidence E is a positive test result, and the hypotheses are H_1 = the disease is present, and H_2 = the disease is absent. The error characteristics of the test are as follows:

$\Pr(E \mid H_1) = 0.95$	$\Pr({\sim}E \mid H_1) = 0.05$
$\Pr(E \mid H_2) = 0.02$	$\Pr({\sim}E \mid H_2) = 0.98$

As Royall explains, $r(H_1, E)$ is sensitive to the prior probability of H_1. Royall derives the following formula for $r(H_1, E)$ in terms of the prior $\Pr(H_1)$ and the likelihood ratio $l(H_1, E)$:

$$r(H_1, E) = \frac{\Pr(H_1 \mid E)}{\Pr(H_1)} = \frac{\Pr(E \mid H_1)}{\Pr(E)} = \frac{l(H_1, E)}{l(H_1, E)\Pr(H_1) + (1 - \Pr(H_1))}$$

In this example, $l(H_1, E) = \frac{\Pr(E \mid H_1)}{\Pr(E \mid {\sim}H_1)} = \frac{0.95}{0.02} = 47.5$. So,

$$r(H_1, E) = \frac{47.5}{47.5 \cdot \Pr(H_1) + (1 - \Pr(H_1))}$$

Thus, $r(H_1, E)$ will be only slightly greater than one in this example, if $\Pr(H_1)$ is sufficiently large. Royall thinks this is unintuitive. He thinks that E should "strongly" favor of H_1 over H_2 in this example, *independently* of the prior probability of H_1 (or H_2).

This first critique of Royall's is rather weak, for several reasons. First, Royall is assuming much more than (LL) here. He is assuming something like the following quantitative generalization of (LL):

(LL$^+$) E favors H_1 over H_2 iff $\Pr(E \mid H_1) > \Pr(E \mid H_2)$, *and* the *degree* to which E favors H_1 over H_2 is given by the *ratio* $\frac{\Pr(E \mid H_1)}{\Pr(E \mid H_2)}$.

What's good for the goose is good for the gander. That is, what prevents the ($\dagger_r$)-Bayesian from generalizing *their* account in a like manner?

[13] Royall seems to assume that Bayesians have just one theory of favoring, based on r. He wrongly attributes such a theory to Carnap. In fact, Carnap explicitly criticizes r (which was used by J.M. Keynes and his teacher E.W. Johnson before him) as a measure of relevance confirmation. Carnap suggests d (and perhaps a family of covariance measures) as a measure of relevance confirmation. So, ($\dagger_d$) would be a much more charitable and accurate relevance-based theory to attribute to Carnap. Royall's worries also apply to ($\dagger_d$), so this is not a fatal historical error.

($\dagger_r^+$) E favors H_1 over H_2 iff $r(H_1, E) > r(H_2, E)$, *and* the *degree* to which E favors H_1 over H_2 is given by the *ratio* $\frac{r(H_1,E)}{r(H_2,E)}$.

It is easy to show that ($\dagger_r^+$) is *logically equivalent to* (LL$^+$).[14] So, it should seem mysterious as to why Royall takes this to be a *critique* of the ($\dagger_r$) account of favoring. On the contrary, it should appear that Royall's account is (to put it mildly) quite consistent with ($\dagger_r^+$).

Second, even if we follow Royall and assume that (in this example) $\mathfrak{c}(H_1, E)$ [and not some function of $\mathfrak{c}(H_1, E)$ and $\mathfrak{c}(H_2, E)$] is the proper Bayesian $\mathfrak{c}$-based measure of the degree to which E favors H_1 over H_2 (as I will explain below, this is not as crazy as it sounds *in this case*), there is no reason that r must be adopted as *the* Bayesian measure of confirmation in this sense. It is true that several contemporary Bayesians [*e.g.*, Milne (Milne, 1996) and Howson & Urbach (Howson and Urbach, 1993)] defend r as the proper measure of confirmation in this sense. But, it is easy to see that ($\dagger_l$) has exactly the property Royall wants in this example. If we take $l(H_1, E)$ [and not some function of $l(H_1, E)$ and $l(H_2, E)$] as the Bayesian measure of the degree to which E favors H_1 over H_2 (in this example), then we obtain the result that the degree to which E favors H_1 over H_2 in this example is 47.5, which is "large" by Royall's own lights. Indeed, even if we take some function of $l(H_1, E)$ and $l(H_2, E)$ as the Bayesian measure of "degree of favoring," then (intuitively) we will also get a "large" value [and one which does not depend on $\Pr(H_1)$]. So, on either kind of quantitative generalization of the Bayesian account of favoring, using l leads to a Bayesian account that agrees with Royall's own theory in this example.[15]

Third, and lastly, it is quite strange for a contrastive empiricist like Royall to make such heavy weather of *this* example. In this example, H_1 is just the logical opposite of H_2. It is easy to see that, in such cases, we will *always* have $\Pr(E \mid H_1) > \Pr(E \mid H_2)$ iff $\mathfrak{c}(H_1, E) > \mathfrak{c}(H_2, E)$, for *any* Bayesian relevance measure $\mathfrak{c}$ of degree of non-relational confirmation.[16] So, looking at examples in which H_1 is equivalent to $\sim H_2$ is *never* going to provide an adjudication between Likelihoodism and Bayesianism. What we need are examples in which H_1 and H_2 are *not* logical opposites. This brings us to Royall's second critique.

[14] This requires only a trivial modification of the proof in footnote 8, above.

[15] What's more, there are compelling reasons to favor l over r, from a Bayesian point of view. See footnotes 8, 22, and 25 for further discussion of this issue.

[16] If $H_1 \dashv\vDash \sim H_2$, then $\Pr(E \mid H_1) > \Pr(E \mid H_2)$ iff $\Pr(E \mid H_1) > \Pr(E \mid \sim H_1)$, which is true iff $\mathfrak{c}(H_1, E) > \mathfrak{c}(\sim H_1, E)$, for any relevance measure you like. Without loss of generality, consider the difference measure, which is such that $d(H_1, E) > d(\sim H_1, E)$ iff $\Pr(H_1 \mid E) - \Pr(H_1) > 1 - \Pr(H_1 \mid E) - 1 + \Pr(H_1)$ iff $2 \cdot [\Pr(H_1 \mid E) - \Pr(H_1)] > 0$ iff E and H are correlated under Pr iff $\Pr(E \mid H_1) > \Pr(E \mid \sim H_1)$.

In his second critique of Bayesian favoring, Royall considers a class of examples in which the space of hypotheses consists of a set of three mutually exclusive and exhaustive propositions $\{H_1, H_2, H_3\}$. Royall correctly points out that, according to Bayesian accounts of favoring, whether E favors H_1 over H_2 in such cases will (sometimes) depend on the prior probability distribution over the entire space of hypotheses. So, in particular, there will be some cases of this kind in which E favors H_1 over H_2 for some values of $\Pr(H_3)$, but not for others. Royall seems to think this is strange, because it implies that a relation between E, H_1 and H_2 depends on some non-relational property of H_3, where H_3 doesn't even participate in the favoring relation in question. Royall does not give specific examples here, so it's not entirely clear what he has in mind here. I have three remarks about Royall's second critique.

First, on a positive note, at least Royall moves away here from cases in which H_1 and H_2 are logical opposites, which, as we saw above, will never be illuminating. So, he's on the right track toward isolating examples that are *capable of* providing crucial contrastive evidence regarding Likelihoodist *vs* Bayesian theories of favoring.

Second, and more negatively, Royall fails to mention that (LL) appears to be equivalent to $(\dagger_r)$.[17] That is, Royall's theory appears *itself* to be a variety of Bayesian favoring. So, how can Royall's theory be any less sensitive to probability distributions over the hypothesis space than the Bayesian $(\dagger_r)$? Of course, it cannot be, unless the (LL) is formulated so as to ensure that the likelihoods that appear in the (LL) are never sensitive to marginal or prior probability distributions over the hypothesis space. Indeed, this is what Likelihoodists try to do. Typically, Likelihoodists use only probability models that determine likelihoods, but neither priors nor posteriors of hypotheses. Bayesians, on the other hand, use complete probability models in which likelihoods, priors, and posteriors are intimately connected by Bayes's Theorem. [This also explains Sober's "unknown priors" *ceteris paribus* clause.] Thus, it is only when we are working with an impoverished "Likelihood-Only" probability model that Likelihoodism is *truly* distinct from Bayesianism. But, in such cases, we cannot *contrast* Likelihoodism and Bayesianism, since Bayesianism will not say anything at all!

This places Likelihoodists in an uncomfortable dilemma. Either they stick to the "pure" form of Likelihoodism, which only operates in exactly the cases that Bayesian theories of favoring are silent, or they allow Likelihoodism to be applied to cases where we have complete prob-

[17] To be fair, Royall's understanding of "likelihood" is sometimes not as a conditional probability at all, but as an *un*conditional probability *entailed by* a hypothesis. As I explain in footnote 2 above, this reading of Royall makes the dilemma below *even worse*. That's why I'm not adopting such a reading of Royall here.

ability models (at least, complete enough for the salient Bayesianism alternatives to render favoring judgments). If Likelihoodists embrace the first horn of the dilemma, then they cannot – by their own lights – provide evidence that *favors* their theory of favoring over Bayesian theories, since Bayesianism will be silent in all such cases. And, if they embrace the second horn of the dilemma, then their theory just *is* a form of Bayesianism, and they must face whatever general problems Bayesian *qua* Bayesian accounts of favoring face, including (some of) Royall's own examples, and examples like mine (and Leeds'), which seem to show that theirs is an *inadequate* form of Bayesianism at that. Moreover, the Likelihoodist's task is made that much more difficult by the fact that the only (non-trivial) thing all Bayesian theories of favoring seem to have in common is (WLL), which is a logical consequence of their own view. In other words, embracing the second horn of the dilemma forces the Likelihoodist to get down in the mud and argue with Bayesians about the precise form that Bayesian confirmation theory itself should take. I see no way that the Likelihoodist can in good conscience embrace the first horn of this dilemma, unless they are willing to concede that the dispute between themselves and Bayesians is not amenable to rational adjudication, in principle (since, in that case, no evidence could ever favor their theory over Bayesian theory). So, it seems to me that the only hope for Likelihoodists is to describe examples involving *complete* probability models (at least, complete enough for contrastive purposes) in which (LL) [hence, $(\dagger_r)$] does an intuitively *better job* of capturing the favoring relations than *competing* Bayesian approaches like $(\dagger_d)$ and $(\dagger_l)$ do. Unfortunately, this argumentative burden has not been met by Likelihoodists. Indeed, most Likelihoodists do not even seem to be aware that such a burden exists. To this end, Royall's second class of examples is the right place to look for crucial thought experiments.

Interestingly, there is a well known problem of this kind in the literature that happens to have just the right structure for the purpose at hand: The Monty Hall Problem (MHP). The (MHP) is as follows:

> Imagine you are on a game show. You are faced with three doors (1, 2, and 3), behind one of which is a prize and behind the other two is no prize. In the first stage of the game, you tentatively select a door (this is your initial guess as to where the prize is). To fix our ideas, letÕs say you tentatively choose door 3 (H_3). Then the host, Monty Hall, who knows where the prize is, opens one of the two remaining doors. Monty Hall can never open either the door that has the prize or the door that you tentatively choose; he must open one remaining door that does not contain the prize. Now you learn (E) that Monty Hall has opened door 1.

In this set-up, H_i = the prize is behind door i, E = Monty opens door #1, and your initial guess is H_3. The issue that is usually addressed concerning the (MHP) is whether the posterior probability of H_2 is greater than the posterior probability of H_3, given E. That is, the standard question is whether $\Pr(H_2 \mid E) > \Pr(H_3 \mid E)$. There is much controversy about this question in the literature. In fact, this question cannot be answered unless further information about the priors $\Pr(H_i)$ and the likelihood $\Pr(E \mid H_3)$ are given (these are not specified in the problem description). But, what about the favoring question: Does E favor H_2 over H_3? Well, of course, this depends on one's theory of favoring. If one has a Likelihoodist [*i.e.*, ($\dagger_r$)] theory of favoring, then the answer is already known, and does not depend on the priors $\Pr(H_i)$ or the likelihood $\Pr(E \mid H_3)$ [except the very weak assumption that $\Pr(E \mid H_3) < 1$], since $\Pr(E \mid H_2) = 1 > \Pr(E \mid H_3)$.[18] Thus, according to Likelihoodism, E favors H_2 over H_3, and this does not depend on the marginal distribution over the hypothesis space (as Royall seems to want). But, if one is a (non-Likelihoodist) Bayesian, then (in general) the answer will depend — to some extent — on the marginals $\Pr(H_i)$ [given a fixed value of the likelihood $\Pr(E \mid H_3)$]. However, the extent to which the answer depends on the priors [given a fixed value of the likelihood $\Pr(E \mid H_3)$], varies from one Bayesian theory to another. If we fix the likelihood $\Pr(E \mid H_3) = \frac{1}{2}$, which is somewhat typical in traditional discussions of the (MHP), then we can analytically determine the set of prior probability distributions relative to which the various Bayesian theories of favoring render the judgment that E favors H_2 over H_3. The best way to see how this analysis pans out is by examining Figure 1.[19]

The theory that is most sensitive to the marginal distribution is the theory based on the posteriors (‡). Only marginal distributions above the line [including the large dot, which represents the equiprobable marginal distribution, often assumed in traditional discussions of the (MHP)] are consistent with $\Pr(H_2 \mid E) > \Pr(H_3 \mid E)$. The next most prior-sensitive favoring theory is the (WLL), which is the top curve on the plot. Only marginal distributions above this top curve are consistent with $\Pr(E \mid H_2) > \Pr(E \mid H_3)$ *and* $\Pr(E \mid {\sim}H_2) \leq \Pr(E \mid {\sim}H_3)$. The next most prior-sensitive theory is ($\dagger_d$), and the least prior-sensitive (non-Likelihoodist!) Bayesian theory is ($\dagger_l$). This ordering of Bayesian theories of favoring in terms of their prior-sensitivity is invariant under

[18] See (Bradley and Fitelson, 2003) for further discussion of the Monty Hall Problem, from both a posterior-probabilistic and a Likelihoodist perspective.
[19] A *Mathematica* notebook containing analytical solutions and further plots for the (MHP) can be downloaded from `http://fitelson.org/monty.nb`.

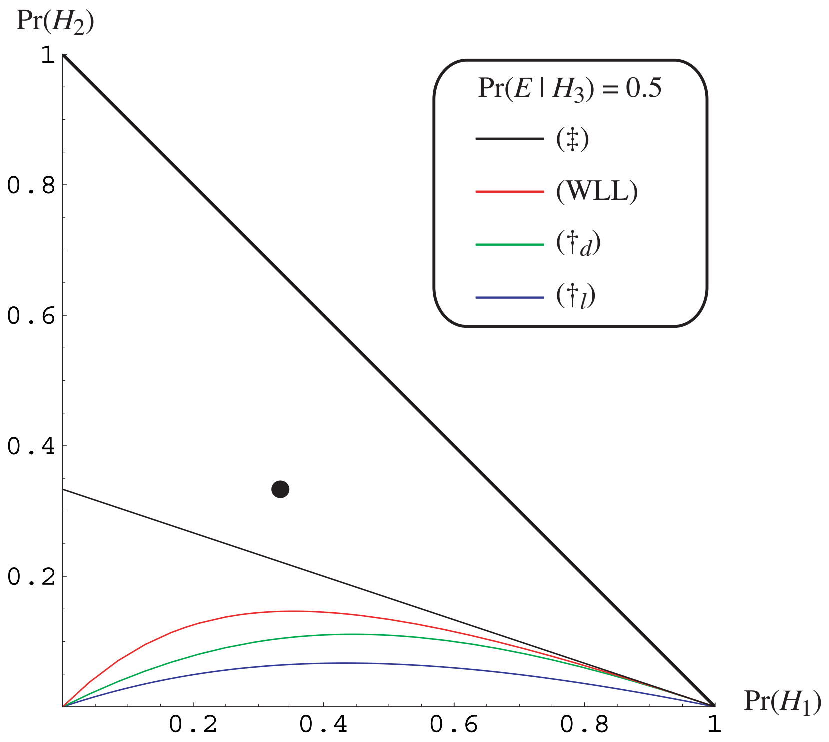

Figure 1. Dependence on Priors of Bayesian Favoring Theories in the (MHP)

changes to the likelihood $\Pr(E \mid H_3)$.[20] More precisely, if $\{T\}$ is the set of marginal distributions over the H_i such that favoring theory T renders the judgment that E favors H_2 over H_3 in the (MHP), then we have the following ordering of inclusion relations:

$$\{\ddagger\} \subset \{\text{WLL}\} \subset \{\dagger_d\} \subset \{\dagger_l\} \subset \{\text{LL}\} = \{\dagger_r\}$$

This gives us a concrete instance of the class of examples Royall seems to have in mind in his second critique of Bayesian favoring. The (MHP) is an example in which Likelihoodism says that the favoring relation

[20] While the *ordering* of theories according to their prior-sensitivity is invariant under changes to $\Pr(E \mid H_3)$, the *quantitative* behavior of each theory's sensitivity to priors, of course, is not. As $\Pr(E \mid H_3)$ approaches zero, the dependence on priors vanishes for all Bayesian theories of favoring (*i.e.*, the class of priors for which E favors H_2 over H_3 converges to the entire simplex), and as $\Pr(E \mid H_3)$ approaches 1, the dependence on priors converges to a maximal value (*i.e.*, the class of priors for which E favors H_2 over H_3 converges to the "small" part of the simplex above the line $\Pr(H_2) = -\frac{1}{2} \cdot \Pr(H_1) + \frac{1}{2}$, which includes the flat marginal distribution — represented by the big dot in Figure 1 — at which $\Pr(H_1) = \Pr(H_2) = \Pr(H_3) = \frac{1}{3}$).

should not depend *at all* on the marginal distribution over the hypothesis space (as Royall seems to want).[21] On the other hand, the various non-Likelihoodist Bayesian theories of favoring will say that the favoring relation in the (MHP) depends *to some extent* on the marginal distribution over the hypothesis space. However, the extent to which (non-Likelihoodist) Bayesianism is prior-sensitive in this sense varies greatly from one Bayesian theory to another, with the least amount of prior-sensitivity being exhibited by the likelihood-ratio-based Bayesian theory: ($\dagger_l$), and the greatest degree of prior-sensitivity being exhibited by the posterior-based account (‡), which has already been shown inadequate on independent grounds. What Royall needs to argue is that the *complete* prior-insensitivity exhibited by (LL) and ($\dagger_r$) in examples like the (MHP) is to be preferred to the *degree* of prior-sensitivity exhibited by the best Bayesian alternative. For various reasons, (which I won't go into here[22]) I think the best Bayesian theory of favoring is ($\dagger_l$), which happens also to be the *least* prior-sensitive theory of the Bayesian alternatives in the (MHP). I will not try to argue here that Bayesians should prefer ($\dagger_l$) to ($\dagger_r$), or that ($\dagger_l$) is preferable to Likelihoodism. My only aim here is to clarify the Likelihoodist *vs* Bayesian dialectic, by sharpening the argumentative burden of the Likelihoodist. If Likelihoodists want to critique Bayesian favoring theory, they will need to focus on examples like the (MHP), which (given a sufficiently rich

[21] Note: There will be many examples in which (LL) and ($\dagger_r$) *do* depend on the marginal distribution, since (in general) Likelihoods – in complete probability models – *also* depend on priors. After all, by Bayes's Theorem, $\Pr(E \mid H_1) = \frac{\Pr(H_1 \mid E)\cdot\Pr(E)}{\Pr(H_1)}$. So, this Likelihoodist talk of "insensitivity to marginal distribution" is rather misleading, assuming that we are operating in a setting where both Bayesian and Likelihoodist theories of favoring can render judgments (which is required, *if* they're to be *contrasted*). What's really at issue, then, is not dependencies on priors *simpliciter* (since *all* theories of favoring have *some* of *that* in the salient contexts), but *kinds* of dependencies on priors. Likelihoodists need to argue that *their* ($\dagger_r$) flavor of prior dependence is preferable to *other competing* flavors, *e.g.*, ($\dagger_d$)'s and ($\dagger_l$)'s.

[22] It is not the purpose of this paper to argue in favor of any particular Bayesian theory of favoring ($\dagger_{\mathfrak{c}}$). From a Bayesian point of view, this battle needs to be fought at the more primitive level of non-relational confirmation. Thus, an argument for ($\dagger_{\mathfrak{c}}$) is an argument for $\mathfrak{c}$. To this end, see (Eells and Fitelson, 2002), (Fitelson, 2001a), (Fitelson, 2001b), (Fitelson, 2002), (Fitelson, 2005), and (Good, 1985) for various reasons to prefer l over both d and r, as the proper Bayesian measure of non-relational confirmation. See footnotes 8 and 25 for more on this Bayesian controversy. It is important to note that not everyone thinks there is an underlying dispute here. For instance, Joyce (2004b) argues that there is no real disagreement at all between ($\dagger_r$) and ($\dagger_l$). If Joyce is right, then this only serves to bolster my dilemma for Likelihoodists, since Likelihoodists would then need to establish something *false* to meet their argumentative burden of providing examples that favor their theory over Bayesianism. So, even if you disagree with me about there being a genuine underlying intramural Bayesian dispute, my dilemma for Likelihoodism still stands.

probability model of the set-up) are at least *capable of* undergirding an adjudication. Moreover, Likelihoodists should stop talking about (‡), which is not a viable Bayesian theory of favoring, and focus on more interesting competitors, like ($\dagger_l$), which are non-trivial to unfavorably contrast against the (LL) and ($\dagger_r$). In the next section, I will propose an alternative, non-Bayesian (partial) approach to favoring, which is (sort of) a "middle way" between Likelihoodism and Bayesianism.

5. Toward an Alternative Theory of Favoring

Principles (*) and (WLL) each identify intuitively plausible sufficient conditions for "E favors H_1 over H_2". (*) rests only on a logical asymmetry, and (WLL) rests only on the following two inequalities:

1. $\Pr(E \mid H_1) > \Pr(E \mid H_2)$, and
2. $\Pr(E \mid {\sim}H_1) \leq \Pr(E \mid {\sim}H_2)$

As we have seen, Likelihoodists cannot, *in general*, accept (*) as a sufficient condition for favoring, since there are cases in which (*) contradicts the (LL). We have also seen that the (WLL) is a logical consequence of the (LL). Thus, the (WLL) is not something the Likelihoodist can reject. Interestingly, there is an intimate relationship between (*) and the (WLL). If the antecedent of (*) is satisfied (*i.e.*, if $E \vDash H_1$, but $E \nvDash H_2$), then inequality (2) will also be satisfied. These considerations can be used to formulate a (partial) alternative account of the favoring relation that should be amenable to both Bayesians and Likelihoodists. The key to reconciliation here is inequality (2), which is the ultimate source of our Bayes/non-Bayes controversy.

Likelihoodists often criticize Bayesians on the grounds that their favoring relations typically depend sensitively on *priors*, which are taken by Likelihoodists to be "subjective" or, at least, lacking in probative value. The preceding considerations show us that this so-called "problem of priors" (PP) is not the essential problem for theories of favoring that go beyond mere likelihoods. The essential problem, and, it seems to me, the essential issue between Likelihoodists and Bayesians is *not* the problem of priors, but the *problem of catch-alls* (PCA). For Bayesians, the (PCA) can be reduced to the (PP), since Bayesians take catch-alls to be reducible to a function of likelihoods and priors. For Bayesians, the catch-call $\Pr(E \mid {\sim}H_i)$ is just a weighted average of the likelihoods of the alternatives to H_i in some partition $\{H_j\}$ of ${\sim}H_i$:

$$\Pr(E \mid {\sim}H_i) = \frac{\sum_{j \neq i} \Pr(E \mid H_j) \cdot \Pr(H_j)}{\sum_{j \neq i} \Pr(H_j)}$$

Because catch-alls are reducible to priors (given likelihoods), there is an important asymmetry between priors and catch-alls (holding fixed knowledge of likelihoods). Assuming that all likelihoods are known, specifying the marginal distribution over the H_i determines the catch-alls. However, the converse is not true. That is, assuming that all likelihoods are known, specifying all catch-alls does not determine the priors. As such, less information is needed to determine catch-alls than to determine priors (given knowledge of likelihoods). Or, to put things another way, you can know all likelihoods and all catch-alls without knowing all priors, but you cannot know all likelihoods and all priors without knowing all catch-alls. So, from a Bayesian point of view, solving the (PP) would automatically solve the (PCA), but solving the (PCA) would not automatically solve the (PP).

This opens the door to a "middle way" between Bayesian and Likelihoodist theories of favoring. Let's say we agree (*arguendo*) with the Likelihoodist about the (PP), and we don't want a theory of favoring that requires us to determine priors (in general). If we are willing to settle for a sufficient condition for favoring (*i.e.*, a partial theory of favoring[23]), then we can formulate such a condition without even mentioning unconditional probabilities. Indeed, using the (WLL) and (*), we can motivate a sufficient condition based only on *comparisons* of conditional probabilities. In fact, we can push things even farther than this, and formulate a sufficient condition based only on comparisons of conditional plausibilities (Friedman and Halpern, 1995) or ordinal ranking functions (Spohn, 1990). In other words, we don't even need quantitative probability theory at all to give a robust sufficient condition for favoring. All we need to know is whether (1) E is more plausible, given H_1 than H_2, and (2) E is no more plausible given $\sim H_1$ than $\sim H_2$. That doesn't *require* any unconditional probability judgments or even any unconditional plausibility comparisons. It seems to me that this is the proper way to address the Likelihoodists worries about priors. This (partial) approach to favoring does not require any knowledge of priors, and so the (PP) is no longer a relevant consideration.

Of course, I expect the following objection at this point: "You've just replaced one intractable problem (PP), with another intractable

[23] Alternatively, we may be willing to settle for a *quasi*-favoring-ordering of hypotheses (by evidence). On such an approach, some pairs of hypotheses would be *incommensurable* in terms of how they are contrastively supported by E (in such cases, we would have *neither* favoring *nor* confirmational neutrality). While principles (*) and (WLL) provide *sufficient* conditions for favoring, there may be no *complete* set of sufficient conditions that would also circumscribe a *necessary* condition for favoring. If so, then the favored-over-by-E relation would only provide a *quasi*-ordering over $\{H_i\}$. This is analogous to the strategy employed by Bovens & Hartmann (2003) in their theory of quasi-coherence orderings of information sets.

problem (PCA). So, how does this really help?" But, this objection is too fast. Our discussion of (*) shows that there are examples in which the logical asymmetries alone allow us to get an *a priori* handle on the salient conditional plausibility judgment. If $E \vDash H_1$ and $E \nvDash H_2$ (*i.e.*, if the antecedent of (*) holds), then it is reasonable *a priori* to hold that $\Pr(E \mid {\sim}H_1) \leq \Pr(E \mid {\sim}H_2)$, and, provided that the example is *also* such that $\Pr(E \mid H_1) > \Pr(E \mid H_2)$, we will be in a position to know the salient comparative claims without knowing anything about priors. And, in such cases, we clearly *should* say that E favors H_1 over H_2, whether we are Likelihoodists or not. What this shows is that the kind of (partial) theory of favoring I have sketched above is not without probative value, and it is also not "subjective" (at least, not in any bad sense of that term). Moreover, it should be a theory that is perfectly acceptable to Likelihoodists and Bayesians alike. Granted, it is a more conservative theory than Likelihoodism, and it is less quantitative than Bayesianism [or quantitative extensions of the (LL), like (LL^+)], but it is not rendered otiose merely by the fact that it appeals to comparisons like (2). What Likelihoodists need to argue is that inequalities like (2) are not *relevant* to favoring claims. It won't do to argue that inequalities like (2) are always "subjective" or without probative value, since in some cases they clearly are not, and not all cases of this kind are cases in which we have any access to prior or unconditional probabilities. So, the standard Likelihoodist line about the (PP) doesn't help with their argumentative burden here.[24] What Likelihoodists need – by their own lights – is evidence that favors Likelihoodism over this alternative, non-Bayesian approach. As before, my present aim is not to argue that this alternative theory of favoring is preferable to Likelihoodism. I only want to point out that Likelihoodists have not met their argumentative burdens for critiquing plausible alternative accounts of favoring.

6. Conclusion

The main aim of this paper has not been to endorse any particular theory of relational support, but only to clarify what's at issue between Likelihoodists and non-Likelihoodists in this connection, and to argue

[24] It will also do no good to argue that the present approach is not a genuine *alternative* to Likelihoodism, since (*i*) the (WLL) is a logical consequence of the (LL), and (*ii*) in the most intuitive cases, the antecedents of both (LL) and (WLL) are true. This response is not open to Likelihoodists because they take pride in the fact that their theory of favoring can be applied to pairs of hypotheses which are logically dependent. See (Forster and Sober, 2004) for a discussion of this advantage of Likelihoodism over (‡), in the context of selecting among nested statistical models.

that Likelihoodists have more philosophical work to do than one might have thought. Specifically, Likelihoodists have not provided concrete evidence that favors their theory of favoring over promising alternative Bayesian theories like ($\dagger_l$). In order to do so, Likelihoodists need to motivate their position on examples like (MHP), which (*i*) involve sufficiently rich probability models (otherwise, the Bayesian alternatives are silent, and can't be contrasted against anything), and (*ii*) are such that Bayesian theories like ($\dagger_l$) *contradict* Likelihoodism [*i.e.*, ($\dagger_r$)]. Moreover, Likelihoodists have not responded to the kinds of examples that seem to favor Bayesian theories like ($\dagger_l$) over Likelihoodism [*i.e.*, ($\dagger_r$)]. In both Leeds' example and in my example involving logical asymmetries, ($\dagger_l$) renders a *prima facie* plausible favoring judgment that contradicts Likelihoodism [*i.e.*, ($\dagger_r$)].[25] As such, these examples provide *prima facie* contrastive evidence in favor of Bayesianism *vs* Likelihoodism. If Likelihoodists are to make a compelling case in favor of their view in contrast to Bayesianism, they need *both* to respond to such examples, *and* to motivate concrete examples of their own [*e.g.*, the (MHP)], which have the requisite contrastive structure. Finally, Likelihoodists have not even addressed alternative, non-Bayesian theories of favoring – like the partial account sketched above, based only on conditional plausibility/likelihood comparisons and principles (*) and (WLL) – which make no appeal whatsoever to priors.

References

Bovens, L. and S. Hartmann: 2003, *Bayesian Epistemology*. Oxford: Oxford University Press.

Bradley, D. and B. Fitelson: 2003, 'Monty Hall, Doomsday and confirmation'. *Analysis* **63**(1), 23–31.

Carnap, R.: 1962, *Logical Foundations of Probability*. Chicago: University of Chicago Press, second edition.

Edwards, A. W. F.: 1992, *Likelihood*. Baltimore, MD: Johns Hopkins University Press, expanded edition. Revised reprint of the 1972 original.

Eells, E. and B. Fitelson: 2002, 'Symmetries and Asymmetries in Evidential Support'. *Philosophical Studies* **107**(2), 129–142.

Fitelson, B.: 1999, 'The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity'. *Philosophy of Science* **66**, S362–S378.

[25] It is worth noting in this connection that ($\dagger_l$) is the *only* relevance-based Bayesian account of relational support we have considered that obeys Principle (*). [Of course, (‡) also obeys (*), but (‡) is inadequate on independent grounds.] This is the underlying reason why ($\dagger_l$) contradicts ($\dagger_r$) in my counterexample to the (LL). And, this, in turn, can be explained by the fact that the measure l is the only relevance measure of non-relational confirmation we have seen that properly handles cases of *conclusive evidence*. See (Fitelson, 2005) for in-depth discussion.

Fitelson, B.: 2001a, 'A Bayesian account of independent evidence with applications'. *Philosophy of Science* **68**(3, supplement), S123–S140.

Fitelson, B.: 2001b, 'Studies in Bayesian Confirmation Theory'. Ph.D. thesis, University of Wisconsin–Madison (Philosophy). The thesis can be downloaded from http://fitelson.org/thesis.pdf.

Fitelson, B.: 2002, 'Putting the irrelevance back into the problem of irrelevant conjunction'. *Philosophy of Science* **69**(4), 611–622.

Fitelson, B.: 2005, 'Inductive Logic'. In: S. Sarkar and J. Pfeifer (eds.): *The Philosophy of Science: An Encyclopedia.* London: Routledge Press.

Forster, M. and E. Sober: 2004, 'Why Likelihood?'. In: M. Taper and S. Lele (eds.): *The Nature of Scientific Evidence.* Chicago: University of Chicago Press.

Friedman, N. and J. Y. Halpern: 1995, 'Plausibility measures: a user's guide'. In: *Uncertainty in artificial intelligence (Montreal, PQ, 1995).* San Francisco, CA: Morgan Kaufmann, pp. 175–184.

Good, I.: 1985, 'Weight of evidence: a brief survey'. In: *Bayesian Statistics, 2* (Valencia, 1983). Amsterdam: North-Holland, pp. 249–269.

Howson, C. and P. Urbach: 1993, *Scientific Reasoning: The Bayesian Approach.* La Salle: Open Court.

Joyce, J.: 2004a, 'Bayes's Theorem'. In: E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy (Summer 2004 Edition).* URL = http://plato.stanford.edu/archives/sum2004/entries/bayes-theorem/.

Joyce, J.: 2004b, 'On the Plurality of Probabilist Measures of Evidential Relevance'. Unpublished manuscript (presented at PSA 2004 symposium on inductive logic).

Leeds, S.: 2000, 'Other Minds, Support, and Likelihoods'. Unpublished manuscript.

Milne, P.: 1996, 'log$[P(h/eb)/P(h/b)]$ is the One True Measure of Confirmation'. *Philosophy of Science* **63**, 21–26.

Popper, K.: 1954, 'Degree of Confirmation'. *The British Journal for the Philosophy of Science* **5**, 143–149.

Royall, R.: 1997, *Statistical Evidence: A Likelihood Paradigm.* London: Chapman & Hall.

Sober, E.: 1994, 'Contrastive Empiricism'. In: *From A Biological Point of View: Essays in Evolutionary Philosophy.* Cambridge: Cambridge University Press, pp. 114–135.

Sober, E.: 2005, 'Is Drift a Serious Alternative to Natural Selection as an Explanation of Complex Adaptive Traits?'. In: A. O'Hear (ed.): *Philosophy, Biology and Life.* Cambridge: Cambridge University Press.

Spohn, W.: 1990, 'A general nonprobabilistic theory of inductive reasoning'. In: *Uncertainty in artificial intelligence, 4*, Vol. 9 of *Mach. Intelligence Pattern Recogn.* Amsterdam: North-Holland, pp. 149–158.

Steel, D.: 2003, 'A Bayesian way to make stopping rules matter'. *Synthese* **58**, 213–227.