The Favoring Relation: The Central Concept in the Debate

- This talk is mainly about claims of the form "E favors H_1 over H_2", where E is some evidential proposition, and H_1 and H_2 are distinct hypotheses, and H_1 and H_2 need not be mutually exclusive (*e.g.*, nested hypotheses allowed).
- As we'll see, this relation is taken as fundamental by Likelihoodists, but it is a derived (*i.e.*, defined) concept for Bayesian confirmation theorists.
- Since Likelihoodists don't think there is any such thing as non-relational support (*i.e.*, confirmation), we must cast this debate as a debate about favoring, so as not to beg any questions against Likelihoodism.
- First, we will be discussing Likelihoodism's account of favoring, and then we'll be comparing and contrasting it with Bayesian accounts of favoring.
- We'll explain the precise relationship between the two accounts, with an eye toward cutting through the rhetoric on both sides. Then, we'll propose a "Middle Way", which will borrow something from each perspective.

Likelihoodism and Favoring: The "Law" of Likelihood I

- According to Likelihoodists (*e.g.*, Royall), the following "Law" captures an essential feature of (*i.e.*, a nec. & suff. condition for) the favoring relation:

(LL) E favors H_1 over H_2 (relative to $\mathcal{M}_L$) iff $\Pr_{\mathcal{M}_L}(E \mid H_1) > \Pr_{\mathcal{M}_L}(E \mid H_2)$.

[$\mathcal{M}_L$ is some "Likelihoodist-Friendly" model, which (usually?!) determines *likelihoods* $\Pr(E \mid H_i)$, but neither *priors* $\Pr(H_i)$ nor *catch-alls* $\Pr(E \mid {\sim}H_i)$. We'll return to this "impoverished" aspect of Likelihoodist models, below.]

- E can be *misleading* favoring evidence for H_1 *vs* H_2 in at least two ways: (*i*) the model $\mathcal{M}_L$ of the experiment could be incorrect, or, even if it is correct, since it is stochastic, (*ii*) H_1 could be false while H_2 is true, even given E.
- It is somewhat misleading to think of $\Pr_{\mathcal{M}_L}$ as a probability function. It isn't, generally, since $\mathcal{M}_L$ (usually?!) does not determine inverses $\Pr(H_i \mid E)$.
- It is somewhat strange that (LL) is considered to be a "Law". There seem to be counterexamples to it. Perhaps it is best thought of as a "*ceteris paribus* law".

Likelihoodism and Favoring: The "Law" of Likelihood II

- There are various sorts of (alleged) counterexamples to (LL). As I will explain below, there are some strong limitations on the kinds of counterexamples that can be generated (even from a Bayesian point of view). My favorite example:
  - A card is to be drawn at random from a standard deck. We assume the standard model $\mathcal{M}$ for well-shuffled decks. Then, let H_1 be the hypothesis that the card is the A♠, E = the card is a ♠, and H_2 = the card is black.
- Intuitively, E favors H_2 over H_1, since $E \vDash H_2$, but $E \nvDash H_1$. However, since $\Pr_{\mathcal{M}}(E \mid H_1) = 1 > \Pr_{\mathcal{M}}(E \mid H_2) = \frac{1}{2}$, the (LL) implies E favors H_1 over H_2.
- Favoring should respect *entailment*. This seems obvious, but (LL) violates it.
- There are other less compelling examples. Steve Leeds takes E = the card is an ace, H_1 = the card is the A♠, and H_2 = the card is the A♠ or the A♣. So, $\Pr_{\mathcal{M}}(E \mid H_1) = 1 = \Pr(E \mid H_2)$, but (Leeds claims) E favors H_2 over H_1.
- Our example trades only on an asymmetry in the *logical relations between E and* H_1, H_2. There is no such relational asymmetry in Leeds' example.

Likelihoodism and Favoring: The "Law" of Likelihood III

- Some Likelihoodists do find such examples compelling. And, basically, they retreat to thinking of (LL) as a "*ceteris paribus* law". Elliott Sober now says:

> The Law of Likelihood has to be restricted to cases in which the probabilities of hypotheses, given observations, are not under consideration (perhaps because they are not known), and one is limited to information about the probability of the observations given different hypotheses.

- This seems like a cop-out. It is one thing to say that (LL) is *true*, and it is another to say that it is "the best we can do" in cases where we happen to have impoverished models. This sounds to me like a *concession* that (LL) is *false*.
- Likelihoodists (by their own "contrastivist" lights!) need to argue that their theory of favoring is *better than alternative theories* that give different answers in some cases. Ideally, they should give examples in which (LL) gives the intuitively right answer, and the alternative theories seem wrong.
- I will argue, below, that *neither* of these argumentative burdens has been met.

Bayesian Theories of Favoring: Preliminaries

- Bayesians are not afraid to posit (and use) *complete* probability models such as our card model $\mathcal{M}$. And, for a Bayesian, favoring is a derived concept – it is defined in terms of *non*-relational support (*viz.*, *confirmation*). Thus,

  (†) E favors H_1 over H_2 (relative to $\mathcal{M}$) iff $\mathfrak{c}_{\mathcal{M}}(H_1, E) > \mathfrak{c}_{\mathcal{M}}(H_2, E)$.

  Here, $\mathfrak{c}_{\mathcal{M}}(H, E)$ is some *relevance measure* of the degree to which E confirms H, relative to a complete probability model $\mathcal{M}$. Here are two such $\mathfrak{c}_{\mathcal{M}}$s:

  – *Ratio*: $r_{\mathcal{M}}(H, E) =_{df} \dfrac{\Pr_{\mathcal{M}}(H \mid E)}{\Pr_{\mathcal{M}}(H)}$

  – *Likelihood-Ratio*: $l_{\mathcal{M}}(H, E) =_{df} \dfrac{\Pr_{\mathcal{M}}(E \mid H)}{\Pr_{\mathcal{M}}(E \mid \sim H)}$
- Many other relevance measures have been proposed, but the contrast between $l_{\mathcal{M}}$ and $r_{\mathcal{M}}$ will be sufficient for our purposes today. Interestingly, $r_{\mathcal{M}}$ has quite strong Likelihoodist tendencies, whereas $l_{\mathcal{M}}$ tends to be more "Bayesian".
- Before we get into all that, we'll discuss Royall's argument against (†) + $r_{\mathcal{M}}$.

Bayesian Theories of Favoring: Royall's Argument Against I

- Let D be the hypothesis that some disease is present. And, let $+$ ($-$) be the proposition that some diagnotic test yields a positive (negative) result. The test procedure in Royall's example has the following error characteristics:

$\Pr(+ \mid D) = 0.95$	$\Pr(- \mid D) = 0.05 = \alpha$ [$\Pr$(Type I error)]
$\Pr(+ \mid \sim D) = 0.02 = \beta$ [$\Pr$(Type II error)]	$\Pr(- \mid \sim D) = 0.98$

- Royall assumes (see below for why) that Bayesians measure the degree to which E confirms H, using the ratio measure $r(H, E) = \frac{\Pr(H \mid E)}{\Pr(H)} = \frac{\Pr(E \mid H)}{\Pr(E)}$.
- He points out (quite correctly) that, in the above example, we will have:

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

  where $l(D, +)$ is just the likelihood ratio $\frac{\Pr(+ \mid D)}{\Pr(+ \mid \sim D)} = \frac{0.95}{0.02} = 47.5$.
- Royall complains that, according to the ratio measure, a positive test result may not provide "strong" evidence for D *vs* $\sim D$ (*e.g.*, if $\Pr(D)$ is "high").

Bayesian Theories of Favoring: Royall's Argument Against II

- I think this argument of Royall's is not compelling, for several reasons:
  – First, as I will show below, Royall *must agree* with (†)+r's *qualitative* favoring judgments, since they *necessarily* accord with the (LL)!
  – So, Royall's objection only makes sense if we assume something *stronger* than the (LL): that the "strength of the favoring relation" (assuming favoring isn't *merely ordinal*!) is proportional to the *ratio* $\frac{\Pr(E \mid H_1)}{\Pr(E \mid H_2)}$.
  – But, *even this* objection can be met by a defender of (†)+r, provided that they define "degree of favoring" as the r-*ratio* $\frac{r(H_1,E)}{r(H_2,E)} = \frac{\Pr(E \mid H_1)}{\Pr(E \mid H_2)}$.
  – Second, if a Bayesian uses l instead of r, Royall's claims are just otiose.
  – What's going on? In this special case where $H_1 = \sim H_2$, we have:

$$\Pr(E \mid H_1) > \Pr(E \mid H_2) \text{ iff } \mathfrak{c}(H_1, E) > \mathfrak{c}(H_2, E).$$

  – Thus, if $H_1 = \sim H_2$, *any* (qualitative) Bayesian theory of favoring *is* (LL)! It's only when H_1 and H_2 are *not* logical opposites that (†) and (LL) can come apart. Indeed, something more general and interesting can be said.

Bayesian Theories of Favoring: The Weak Law of Likelihood I

- Consider the following principle, which is *strictly weaker than* (LL):

(WLL) E favors H_1 over H_2 (relative to $\mathcal{M}$) if [sufficient condition only] $\Pr_{\mathcal{M}}(E \mid H_1) > \Pr_{\mathcal{M}}(E \mid H_2)$ *and* $\Pr_{\mathcal{M}}(E \mid \sim H_1) \leqslant \Pr_{\mathcal{M}}(E \mid \sim H_2)$.

- Bayesians who accept (†) and use l to measure confirmation will, *obviously*, accept (WLL). What about other Bayesian measures of confirmation?

  Theorem (Joyce). *(†) implies (WLL), for* ***any*** *Bayesian relevance measure* $\mathfrak{c}$.
- So, (WLL) will be implied by *any* reductive Bayesian account of favoring, and by Likelihoodism [the counterexamples to (LL) accord with (WLL)!].
- Here is another interesting Theorem (used by Peter Milne to argue *for* r):

  Theorem. *(†) implies (LL), if* $\mathfrak{c} = r$ *[very few other* $\mathfrak{c}$*s have this property].*
- Thus, the qualitative favoring judgments endorsed by Likelihoodists must accord with reductive Bayesian judgments rendered by the (†)+r-theory.
- Next, I'll talk a bit more about (WLL), (LL), and the Bayesianism debate.

Bayesian Theories of Favoring: The Weak Law of Likelihood II

- Likelihoodists are "contrastive epistemologists". They think that we can have evidence that *favors* one theory over another, but *not* evidence that *confirms* a theory (in a non-relational sense). But, do they practice what they preach?
- I think not. By their own lights, Likelihoodists need to provide evidence in favor of their theory of (qualitative) favoring [(LL)] *over* candidate Bayesian *alternative* theories, which, in general, imply only (WLL), but not (LL).
- I prefer ($\dagger_l$)-theory. [Note: ($\dagger_r$) $\vDash$ (LL), and I think (LL) is false. But, I think r and other alternatives to l are inadequate for various independent reasons.]
- So, the Likelihoodist needs to provide arguments in favor of (LL) *vs* ($\dagger_l$). How can this be done? Note: there are clearly no examples in which (LL) is true but (WLL) is false! Are there examples which favor (LL) over ($\dagger_l$)?
- I'm not sure. Certainly, Royall's example is *not* such a case (since the two theories must agree there). But, we *do* have examples that clearly favor ($\dagger_l$) over (LL). I'll discuss a possible example in favor of Likelihoodism later.

Constraints on Possible Counterexamples to (LL)

- It would have been really devastating to (LL) if we could find examples in which, say, E confirms H_1 but disconfimrs (or is independent of) H_2, but according to (LL), E favors H_2 over H_1. Happily, this is impossible.
  - If E confirms H_1, then $\Pr(E \mid H_1) > \Pr(E)$.
  - So, under this assumption, if E disconfirms H_2, then $\Pr(E \mid H_2) < \Pr(E)$, which forces $\Pr(E \mid H_1) > \Pr(E) > \Pr(E \mid H_2)$, and the (LL) is correct.
  - Similarly, under this assumption, if $E \perp\!\!\!\perp H_2$, then $\Pr(E \mid H_2) = \Pr(E)$, which forces $\Pr(E \mid H_1) > \Pr(E) = \Pr(E \mid H_2)$, and the (LL) is correct.
- This is why *any* counterexample to (LL) must be such that either E confirms both H_1 and H_2 or E disconfirms both H_1 and H_2 or E is independent of both H_1 and H_2. In our example and in Leeds' too, E confirms both H_1 and H_2.
- The same holds for ($\dagger_l$), of course, which is not surprising, since ($\dagger_l$)-theory is *based on* (relevance) confirmation. Note: if one defines favoring using *posteriors*, this is no longer true – a *reductio ad absurdum* of *that* theory!

Likelihoodism, Bayesianism, and the Problem of Priors

- The reason Likelihoodists are skeptical about non-relational support (in particular, Bayesian confirmation) is that they are skeptical about the existence of "objective" (or epistemically probative) priors $\Pr_{\mathcal{M}}(H_i)$.
- The problem, according to the Likelihoodist, is that Bayesian models $\mathcal{M}$ must posit certain values for the priors, but these posits have no objective basis, or epistemic probative force. This is "the problem of priors".
- This is why Sober says that the (LL) should be restricted to cases in which priors are not objectively known (or knowable). He doesn't offer a theory of what to do when they *are* known (neither does Royall, of course).
- I want to suggest another direction for this debate. I suggest (taking (WLL) as a point of departure) that what is really at issue is not the probative force of *priors* $\Pr(H_i)$, but the probative force of *catch-alls* $\Pr(E \mid {\sim}H_i)$.
- And, I want to urge that only *ordinal structure* is needed, not cardinality. This leads to a "Middle Way" based on the Weak Law of Likelihood (WLL).

A "Middle Way" Based on the (WLL)

- The (WLL) tells us that in order to show that E favors H_1 over H_2 (relative to $\mathcal{M}$) it is sufficient to establish the following two comparative claims:
  1. $\Pr_{\mathcal{M}}(E \mid H_1) > \Pr_{\mathcal{M}}(E \mid H_2)$
  2. $\Pr_{\mathcal{M}}(E \mid {\sim}H_1) \leqslant \Pr_{\mathcal{M}}(E \mid {\sim}H_2)$
- Therefore, all we *really* need are two pieces of *ordinal ranking* information. Is E more plausible, given H_1 than given H_2? And, is E less plausible given ${\sim}H_1$ than ${\sim}H_2$? No "unconditional plausibility comparisons" are needed.
- So, we don't need quantitative probability theory at all to capture the favoring relation. We could (*e.g.*) use Spohn's ordinal ranking functions. Then, (1) and (2) would become $\kappa(E \mid H_1) > \kappa(E \mid H_2)$ and $\kappa(E \mid {\sim}H_1) \leqslant \kappa(E \mid {\sim}H_2)$.
- This avoids the problem of priors, since we don't need to establish any claims about unconditional probability. But, it highlights what I'll call *the problem of catch-alls.* Claim (2) is the controversial one, and it's about catch-alls.

The Problem of Catch-Alls

- At this point, many will immediately object, as follows:
  - Now, you've just replaced one intractable problem with another!
- Too fast? We've already seen clear-cut examples in which the problem of catch-alls is tractable, even though the corresponding problem of priors isn't.
- As long as $E \vDash H_2$, but $E \nvDash H_1$ (and E, H_1, H_2 are contingent), we will have $\Pr(E \mid {\sim}H_1) > \Pr(E \mid {\sim}H_2)$, and (WLL) won't imply that E favors H_1 over H_2. And, (given regularity) ($\dagger_l$) yields the judgment that E favors H_2 over H_1.
- The intuitive judgment that E favors H_2 over H_1 (and not the other way-round) in such examples is not grounded in any assumptions about priors. It is simply the asymmetry in logical entailments that does the work.
- So, here's a case where we can settle the comparative catch-all question without settling the comparative prior question. The converse can't occur. If we know the ordering of the priors (and likelihoods), then (1) settles (2).
- So, the catch-all problem is (in this sense) "easier". Quantitative example …

A Famous and Illustrative Quantitative Example I

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, lets 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.

- Let H_i = the prize is behind door i, and E = Monty opens door #1. What we will be asking is whether (and under what conditions) E favors H_2 over H_3.
- This depends on one's theory of favoring! If one adopts a Likelihoodist (LL) account, then one *must* say that E favors H_2 over H_3 [this depends only on a *regularity* assumption, since $\Pr(E \mid H_2) = 1$]. What about Bayesian ($\dagger_c$)s?
- This is best visualized with a set of animated plots. What the plots show is an ordering of favoring theories in terms of "degree of dependence on priors".

A Famous and Illustrative Quantitative Example II

- The (LL) [($\dagger_r$)] judgment holds for *all* prior distributions over the hypotheses.
- ($\dagger_l$) says E favors H_2 over H_3 for a smaller set of prior distributions.
- Next is ($\dagger_d$): ($\dagger$)+ the *difference* measure $d(H, E) = \Pr(H \mid E) - \Pr(H)$.
- This is followed by (WLL), which holds for a slightly smaller class of priors.
- (Posteriors) $\Pr(H_2 \mid E) > \Pr(H_3 \mid E)$ holds for the narrowest class of priors.
- Thus, the sets of priors $\{T\}$ for which the various favoring theories (T) say that E favors H_2 over H_3 in the Monty Hall Problem are related as follows:

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

- This is a potentially useful example for arguing *in favor of* (LL). If you have the intuition that E favors H_2 over H_3 here — *independently of* $\Pr(E \mid {\sim}H_2)$ *vs* $\Pr(E \mid {\sim}H_3)$ — then you're leaning toward (LL) [if you're Bayesian, ($\dagger_r$)].
- I would say that the problem is *underspecified*. We need information about $\Pr(E \mid {\sim}H_2)$ *vs* $\Pr(E \mid {\sim}H_3)$, which we aren't given in the problem statement.

Monty Hall & Royall's Second Argument Against Bayesian Favoring

- Royall gives a second argument against Bayesian theories of favoring. This one does not involve a specific example, but a *class* of examples that Royall takes to be counterexamples – examples involving *three or more* hypotheses.
- Royall points out that, according to Likelihoodism (LL), the favoring relation "E favors H_2 over H_3" does not depend on what other hypotheses (*e.g.*, H_1) are included in the space of alternative hypotheses (*viz.*, their priors $\Pr(H_i)$).
- But, according to [non-($\dagger_r$)!] Bayesian ($\dagger_c$) theories of favoring, there will (for $n \geqslant 3$) be sensitivity to the prior distribution over the hypothesis space.
- Monty Hall is a concrete example. Two points: (*i*) it is not the priors that matter, but the relationship between $\Pr(E \mid {\sim}H_2)$, $\Pr(E \mid {\sim}H_3)$, $\Pr(E \mid H_3)$; and, (*ii*) it's not obvious that (LL) [and ($\dagger_r$)!] give the correct answer here.
- According to ($\dagger_l$), E favors H_2 over H_3 iff $\frac{\Pr(E \mid {\sim}H_3)}{\Pr(E \mid {\sim}H_2)} > \Pr(E \mid H_3)$. Thus, if $\Pr(E \mid {\sim}H_2) \gg \Pr(E \mid {\sim}H_3)$, then l will say that E does *not* favor H_2 over H_3. This isn't obviously wrong (and needn't be described in terms of priors)!