Overview of Presentation

- Some Preliminaries
  - Background on Bayesianism (*i.e.*, probabilistic epistemology)
  - Background on Bayesian Confirmation Theory
- The Problem of Old Evidence
  - What's the problem?
  - Two Key Assumptions Behind the Problem
- Two Solutions to the Problem of Old Evidence
  - Solution *Via* Denial of the First Key Assumption ("Easy")
  - Solution *Via* Denial of the Second Key Assumption ("Hard")
- Concluding Remarks & Questions

Preliminaries I: Some General Bayesian Background I

- Bayesianism (*i.e.*, Bayesian *epistemology*) assumes that the degrees of belief (or credence) of rational agents are *probabilities*.
- $\Pr(H)$ denotes the (rational) degree of belief in a proposition H. This is called the *un*conditional (or *prior*) probability of H.
- $\Pr(H \mid E)$ is (roughly) the degree of belief a rational agent would assign to H after learning E *with certainty*. I will call this the *strict posterior probability* of H (on E). This is defined as: $\Pr(H \mid E) =_{df} \frac{\Pr(H \,\&\, E)}{\Pr(E)}$.
- A simple example (to which we will return later): Let H be the hypothesis that a card (drawn at random from a standard deck) is a ♠, and let E be the observation that the card is the ace of ♠. In this example, we have:
- $\Pr(H) = 13/52 = 1/4$ and $\Pr(H \mid E) = \frac{\Pr(H \,\&\, E)}{\Pr(E)} = \frac{\Pr(E)}{\Pr(E)} = 1$. So, in this example, learning E (*with certainty*) *raises* the probability of H.

Preliminaries II: Some General Bayesian Background II

- Let $\Pr^a_{\text{old}}(H)$ be the probability agent a assigns to H *before* E is learned. And, let $\Pr^a_{\text{new}}(H)$ be a's Pr-assignment to H *after* (exactly) E is learned.
- On the orthodox approach, evidence is always learned with certainty. This is because orthodox Bayesianism requires: $\Pr^a_{\text{new}}(H) = \Pr^a_{\text{old}}(H \mid E)$. So, the probability a assigns to E after learning E is just $\Pr^a_{\text{old}}(E \mid E) = 1$.
- Bayesians have ways of computing posterior probabilities even in cases where E is learned *without* certainty. The most popular way is called *Jeffrey Conditionalization*. On this approach, we have the following:

$$\Pr^a_{\text{new}}(H) = \Pr(H \parallel E) = \Pr^a_{\text{old}}(H \mid E) \cdot \Pr^a_{\text{new}}(E) + \Pr^a_{\text{old}}(H \mid \neg E) \cdot \Pr^a_{\text{new}}(\neg E)$$

- When E is learned with certainty, we have $\Pr^a_{\text{new}}(E) = 1$, and $\Pr(H \parallel E) = \Pr(H \mid E)$. For simplicity, I will hereafter write $\Pr(H \parallel E)$ rather than $\Pr_{\text{new}}(H)$, and I will drop the "old" subscripts altogether.

Preliminaries III: Background to Confirmation Theory

- In (orthodox) Bayesian confirmation theory, evidence E *confirms* (or *supports*) a hypothesis H if learning E *raises the probability of* H.
- If learning E *lowers* the probability of H [$\Pr(H \parallel E) < \Pr(H)$], then E *dis*confirms (or *counter*-supports) H, and if learning E does not change the probability of H, then E is confirmationally *irrelevant* to H.
- How should one measure the *degree* to which E confirms H? Various functions of $\Pr(H \parallel E)$ and $\Pr(H)$ have been proposed and defended.
- Among the proposals, we find the following three popular candidates.[a]
  - The *Difference*: $d(H,E) =_{df} \Pr(H \parallel E) - \Pr(H)$
  - The Log-*Ratio*: $r(H,E) =_{df} \log\left[\frac{\Pr(H \parallel E)}{\Pr(H)}\right]$
  - The Log-*Likelihood-Ratio*: $l(H,E) =_{df} \log\left[\frac{\Pr(H \parallel E)\cdot(1-\Pr(H))}{(1-\Pr(H \parallel E))\cdot\Pr(H)}\right]$

[a] Logs of the ratios are taken to ensure (i) that they are +/−/0 when E confirms/disconfirms/is irrelevant to H, and (ii) that they are *additive* in various ways. This is merely a useful convention.

What is the Problem of Old Evidence?

- Glymour was the first to clearly explain the problem (my parens, *etc.*):

  …suppose that evidence E is known (with certainty) at time t, before theory H is introduced. Because E is known (with certainty) at t, $\Pr_t(E) = 1$. Further, because $\Pr_t(E) = 1$, the likelihood of E, given H, $\Pr_t(E \mid H)$ is also 1. We then have (by *BT*):

  $$\Pr_t(H \mid E) = \frac{\Pr_t(H) \cdot \Pr_t(E \mid H)}{\Pr_t(E)} = \Pr_t(H)$$

  The conditional probability of H on E is therefore the same as the prior probability of H: E cannot constitute evidence for H in virtue of the positive relevance criterion…

- Two key assumptions are implicit premises in this argument.

  (*i*) All (rational) Bayesian learning is learning *with certainty*. That is, $\Pr^a_{\text{new}}(E)$ is always 1, and (rational) Bayesian agents a always update *via strict* conditionalization: $\Pr^a(H \parallel E) = \Pr^a(H \mid E)$.

  (*ii*) All Bayesian confirmation judgments (made by rational agents a) are made using the agent's *own up-to-date, personal* probability function $\Pr^a_{\text{new}}$. Together with (1), this implies $\mathfrak{c}^a_{\text{new}}(H, E) = 0$. That is to say, E can neither confirm nor disconfirm H, once a has learned E!

An "Easy" Way Out *Via* Denying Premise (*i*) — I

- Earman suggests, and quickly rejects, an "easy" way out *via* denying (*i*):

  …denying that $\Pr_{\text{new}}(E) = 1$ only serves to trade one version of the old-evidence problem for another. Perhaps it is not certain in November 1915 that the true value of the anomalous advance (of the perihelion of Mercury) was roughly 43″ of arc per century, but most members of the scientific community were pretty darn sure, *e.g.*, $\Pr_{1915}(E) = .999$. Assuming that Einstein's theory does entail E, we find that the confirmatory power $\mathfrak{c}_{1915}(H, E)$ is …less than .0001002. This is counterintuitive …

- A careful reconstruction of Earman's argument reveals that it trades on the following property of the difference measure of confirmatory power d:

  (1) If H entails E, then $\Pr_{\text{new}}(E) \approx 1$ implies $d_{\text{new}}(H, E) \approx 0$.

- This argument has two flaws. First, (1) does hold for d and r, but it does *not* hold for l (contrary to what Earman says in a footnote in *Bayes or Bust?*). Second, this argument only applies to the case of *deductive evidence* ($H \vDash E$). As it turns out, only one of these problems can be fixed…

An "Easy" Way Out *Via* Denying Premise (*i*) — II

- We can fix the second problem, since the following is also a theorem:

  (2) If $\Pr_{\text{new}}(E) \approx 1$, then $d_{\text{new}}(H, E) \approx 0$ and $r_{\text{new}}(H, E) \approx 0$.

- However, the first problem *cannot* be fixed, because:

  (3) *Even if* H entails E and $\Pr_{\text{new}}(E) \approx 1$, $l_{\text{new}}(H, E)$ can be *arbitrarily large*.

- As Earman himself admits, this would not be a "crazy" way out, since:

  There are both historical and philosophical reasons for such a stance …(regarding his Mercury example) …the literature of the period contained everything from 41″ to 45″ of arc per century as the value of the anomalous advance …and even the weaker proposition that the value lies somewhere in this range was challenged by some astronomers. …Bayesians are hardly at a loss here, since Jeffrey has proposed a replacement for strict conditionalization which allows for uncertain learning.

- Indeed, this is precisely my idea for an "easy" way out *via* denying (*i*) …

An "Easy" Way Out *Via* Denying Premise (*i*) — III

- Earman's rejection seems premature. A more charitable reading of the proposed resolution seems only to depend on the following two claims:

  (4) A rational Bayesain agent should not (necessarily) assign probability 1 to contingent empirical propositions (*e.g.*, evidential propositions E).

  (5) The (log) likelihood-ratio measure l is an adequate (even preferable) way for rational Bayesian agents to gauge confirmational power.

- Claim (4) seems quite reasonable. Indeed, this is the cornerstone of Jeffrey's "radical probabilism". Others, including Carnap, have also endorsed principles consistent with (4) (*e.g.*, *regularity* of the credence function Pr).
- Claim (5) is non-trivial. In my dissertation, I have argued that l should be preferred to d (Earman's favorite measure). I.J. Good argues for l too…
- Earman's reaction to this (*via* email): "I like it! It is neat and simple and avoids the contortions required by many of the proposed solutions."

A "Hard" Way Out *Via* Denying Premise (*ii*) — I

- Most Bayesian commentators have accepted premise (*i*) (at least, *arguendo*), and have focused their attacks on premise (*ii*).
- For instance, Howson suggests that, for confirmational purposes, agent a should *not* use her own, up-to-date personal probability $\Pr^{a}_{\text{new}}(\cdot)$. Rather, she should use the personal probability of one of her *counterparts* a'.
- Howson suggests that the relevant counterpart is one just like a, except for not having learned E. But, how do we evaluate the *counterfactual* "had a not learned E, her probability function would have been $\Pr^{a'}_{\text{new}}(\cdot)$"?
- Others propose *historical* approaches, which recommend "turning back the clock" to a time just before E was learned, and noting that a confirmation event *had* occurred. But, then all confirmation judgments are *past tense*.
- I suggest that confirmation judgments are *neither* counterfactual *nor* historical, but *hypothetical* (or *logical* — *i.e.*, they are *not personalistic*).

A "Hard" Way Out *Via* Denying Premise (*ii*) — II

- Let's take a hint from practicing Bayesian statisticians (Bernardo & Smith):

  One further point about the terms prior and posterior is worth emphasizing. *They are not necessarily to be interpreted in a chronological sense*, with the assumption that 'prior' beliefs are specified first and then later modified into 'posterior' beliefs.
- Indeed, when reading Bayesian statistics textbooks (like Bernardo & Smith's), one notices immediately the lack of any personalistic or chronological aspects of the probabilities used in the statistical modeling.
- For instance, we might read something like "A sequence of 1000 Heads in a row in successive tosses of a coin (E) supports a bias in favor of heads (H)."
- When things get rigorous, we find that there is a (complete) *probability model* $\mathcal{M}$ of the coin's behavior, which is such that: $\Pr_{\mathcal{M}}(H \mid E) > \Pr_{\mathcal{M}}(H)$.
- We *never* find claims like "A sequence of 1000 Heads in a row in successive tosses of a coin (E) supports a bias in favor of heads (H) — *unless you have already observed the sequence E, in which case it doesn't*."

A "Hard" Way Out *Via* Denying Premise (*ii*) — III

- Bayesian statisticians rarely, if ever, use non-strict conditionalization. So, they need not deny premise (*i*). Thus, charitable interpretation of Bayesian statistical practice seems to lead us to a denial of premise (*ii*).
- When Bayesian statisticians say "E supports (confirms, favors, *etc.*) H," this is understood to be *relative to a (complete) statistical model* $\mathcal{M}$ — *not* to the background knowledge $\mathcal{K}$ of some agent a (rational or otherwise).
- So, confirmational claims are true (or false) *timelessly*, and in a way that does not depend on what any agent may or may not know (or believe).
- Back to our example (E = card is **A**♠, H = card is ♠). In the standard model $\mathcal{M}$ of this set-up, we have $\Pr_{\mathcal{M}}(H \mid E) > \Pr_{\mathcal{M}}(H)$, and E confirms H (timelessly).
- Similarly, the advance of the perihelion of Mercury (E) *timelessly* favors Einstein's theory (H_1) over Newton's (H_2), since the canonical statistical model $\mathcal{M}$ which pits H_1 *vs* H_2 *vis-à-vis* E is s.t. $l_{\mathcal{M}}(H_1, E) > l_{\mathcal{M}}(H_2, E)$.

A "Hard" Way Out *Via* Denying Premise (*ii*) — IV

- Inductive support is a three-place relation. When we say "E inductively supports H," this is always shorthand for "E inductively supports H *in* $\mathcal{M}$."
- If we are merely making some sort of autobiographical remark, then what we should say is "E raised (or would raise) *my* personal degree of belief in H."
- But, if we are making a claim about *inductive support*, then the third argument is *not* my (historical or counterfactual) state of knowledge, but a (logical or statistical) *probability model* which confers probabilities on H, E, *etc*.
- This proposal has problems of its own. How do we select the right models? Where do the probabilities come from, and how are they to be interpreted?
- I think some progress on the problem of selecting statistical models has been made. But, the problem of "logical" or "quasi-logical" probabilities is notoriously difficult (Carnap worked on this unsuccessfully for 60+ years!). I am optimistic that a compelling story can be told here, but that'll have to wait for another day (James Hawthorne's recent work is along similar lines) ...