Bayesian Confirmation Theory

- Administrative: lots of neat paper topics! ...
- Bayesian Confirmation Theory
  - Some Problems for Bayesian Confirmation Theory
    - "Old evidence", "new theories", and Bayesian confirmation
      - Quick Review of Old-Evidence/New Theories Problem
      - Review of Three traditional "ways out"
      - Toward a new "way out"
    - The (infamous) problem of subjectivity (priors, etc.)
      - Where do priors come from?
      - Where did the objectivity of science go?
      - Some Bayesian responses
  - Time Permitting: Some "Success Stories" for Bayesianism

Old Evidence/New Theories I

- As far as I can tell, the reasoning behind the problems of old-evidence and new-theories is something like the following argument:
  1. A rational Bayesian agent a is *always* (*i.e.*, at all times t) obliged to make *all* judgments in accordance with their "most recent"/"most well-informed" degrees of belief $\Pr_t^a(\cdot)$.
  2. $\Pr_t^a(\cdot)$ is *conditionalized* on *all* of a's evidence up to t (E_t^a).

  $\therefore$ 3. $\Pr_t^a(H \mid E) = \Pr_t^a(H)$, for all H and for all $E \in E_t^a$.
- If (i) a always learns by conditionalization, and (ii) a is always obliged to make all judgments according to their most well-informed degrees of belief $\Pr_t^a(\cdot)$, then $\Pr_t^a(H \mid E) = \Pr_t^a(H)$. That is, a's "old" evidence $E \in E_t^a$ can no longer (*i.e.*, for any $t' > t$) confirm anything!
- Moreover, future (as-yet-unimagined, "new") hypotheses (or theories) will not be confirmable by any evidence observed before their invention.

Old Evidence/New Theories II

- There are *many* proposed resolutions of the old-evidence problem:
  - **Historical**. If we want to re-capture the intuition that a's evidence observed at some past time t (E_t^a) confirms H now, then we should re-think of confirmation as a *historical event*, and exploit the *historical fact* (true *timelessly*!) that $\Pr_t^a(H \mid E_t^a) > \Pr_t^a(H)$.
  - **Counterfactual**. Re-think of confirmation as a *counterfactual event* for a's *counterpart* a' s.t. $\Pr_{now'}^{a'}(H \mid E_t^{a'}) > \Pr_{now'}^{a'}(H)$.
  - **Logical Relation**. Rational Bayesian agents need *not* be *logically omniscient*, and so an agent may learn $H \vdash E$. By assuming various things about how probabilities get assigned to "$H \vdash E$", Garber *et al* show that agents a can have $\Pr_{now}^a(H \mid H \vdash E_t^a) > \Pr_{now}^a(H)$. Earman (chapter 5) has an excellent discussion of this account.
- The logical relation approach (unlike the others) can also cope with the problem of *new theories* (H *not even pondered* before E learned).

Old Evidence/New Theories III

- Aren't statistical inferences really of the form "E favors H_1 over $\{H_2, \ldots, H_n\}$, *relative to an experimental design (and context)* K"?
- Why aren't such (contextualized) claims true in a *timeless* way? And, why should they be *undermined* if E is "currently known" or if a new hypothesis H^* is "currently out there, but as yet undiscovered"?
- This idea is *neither* "historical" *nor* "counterfactual." *Why?*
- **Analogy**: When we say that an argument $\frac{p}{\therefore q}$ is *inductively strong*, we must be clear that we mean *inductively strong* — *relative to background* K. If K' contains information that *undermines* or *defeats* the inference from p to q, then — relative to K' — it may *not* be a strong inference.
- What this means is that inductive strength is *indexical* (or *contextual*). But, why (as B & S might ask) must that render it *chronological*?
- Can you apply this idea to the "old evidence"/"new theory" problem?

Old Evidence/New Theories IV

- An *experimental design* or (perhaps better) an *inferential model* $\mathcal{M}$ specifies (all) probabilistic[a] relations between a set of competing hypotheses $\mathbf{H}$, a set of possible (*i.e.*, generable in $\mathcal{M}$) observations $\mathbf{E}$, and a collection of (perhaps sets of) background conditions $\mathbf{K}$.
- Experimental designs or inferential models $\mathcal{M}$ are *abstract objects* (*i.e.*, they are *neither historical nor counterfactual!*) that we use to frame claims about inductive inferences (or claims about confirmation, *etc.*).
- "E confirms H" := "$E \in \mathbf{E}$ confirms $H \in \mathbf{H}$, relative to $K \in \mathbf{K}$ — *within some experimental design or inference model* $\mathcal{M} = \langle \mathbf{H}, \mathbf{E}, \mathbf{K} \rangle$."
- Such claims are true in a *timeless* (albeit *indexical*) way, that does not depend on what any agent knows (at any time). **Moral**: *The probabilities an agent uses to make decisions are not necessarily the same as those an agent appeals to in an inferential model or experimental design.*

[a] Careful! What *are* these "probabilities"? *Hypothetical* degrees of belief, perhaps?

Problems of Subjectivity I

- The most infamous complaint about Bayesian confirmation is that it is "*subjective*", and so it cannot capture "*objective*" evidential relations.
- The main worry seems to be about the "prior" probabilities of hypotheses $\Pr(H)$. Where do they come from? Can agents just pick priors "at random" if they like? Are there "objective" priors?
- This is important, since judgments of degree of confirmation will depend on the values of the priors (as will the posteriors!).
- Many attempts have been made to find "objectively correct" priors, in cases where "no assumptions" are made ("informationless" priors).
- The *Principle of Indifference* (or *Insufficient Reason*) says that if one has "no relevant information", then one should assign *equal* probabilities to all (mutually exclusive, exhaustive) possibilities.
- PI is *very difficult* to apply consistently. See VF's *Laws and Symmetry*.

Problems of Subjectivity II

- No satisfactory account of "informationless (or objective) priors" has been formulated (see B & S's *Bayesian Theory* and VF's *Laws and Symmetry*).
- Most Bayesians have given up on the "holy grail" of objective or informationless priors. *Merger of opinion* results are a much more popular way of restoring "objectivity" (or "intersubjectivity").
- It can be shown (see Earman's chapter 6) that (given weak assumptions about the process) the opinions of a community of Bayesian agents will "merge in the long run," as more and more data are collected.
- Even if Bayesians begin with radically different priors, these priors will eventually be "washed out" by incoming evidence (*e.g.*, 10^{100} tosses).
- There are problems with these results (see pages 147–149 of Earman).
  - There is no way to tell (generally) how long the "long run" is.
  - What about degrees of support *now*?

"Success Stories" of Bayesian Confirmation I

- It is sometimes claimed (see Earman page 64, sort of tongue-in-cheek) that Bayesian confirmation is able to "winnow a valid kernel of" previous (deductive) accounts of confirmation "from their chaff."
- Earman's cavalcade of "success stories" of Bayesian confirmation is fascinating. This list includes:
  - How Bayesian confirmation nicely generalizes H-D
  - How BC handles the paradoxes of instance confirmation (ravens, etc.)
  - How BC handles "evidential variety" or "diversity"
  - How BC handles the Quine-Duhem problem
  - How BC handles "grue" like paradoxes
- I will discuss each of these, in turn (leaving many details for possible paper topics and exercises — see Earman for many good hints!).

"Success Stories" of Bayesian Confirmation II

- The most famous problem in confirmation is Hempel's paradox of the ravens. Recall this paradox was that both $E = Ra \,\&\, Ba$ and $E' = {\sim}Ba \,\&\, {\sim}Ra$ confirm $H = (\forall x)(Rx \to Bx)$.
- One key difference between Bayesian and deductive accounts of confirmation is that Bayesian accounts are *contextual* (or *three-place*). That is, we must always say "E confirms H, *relative to* K."
- IJ Good argues that, relative to some *conceivable* (but non-actual!) K's, $Ra \,\&\, Ba$ may not even confirm that all ravens are black!
  - $K :=$ "We are in one of two worlds: (w_1) with 100 black ravens, no nonblack ravens, and 1 million other birds, or else (w_2) with 1,000 black ravens, 1 white raven, and 1 million other birds."
  - A bird a is selected at random from all the birds, and it is seen that $Ra \,\&\, Ba$. This seems to *dis*confirm that all ravens are black!

"Success Stories" of Bayesian Confirmation III

- Let's assume we're not in one of Good's pathological worlds, and that black ravens do confirm that all ravens are black.
- The Bayesian strategy is to show that — even if a nonblack nonraven confirms that all ravens are black — it confirms that all ravens are black *much less strongly* than a black raven does.
- $\mathfrak{c}(H, E \mid K) \gg \mathfrak{c}(H, E' \mid K)$, where K is our *actual* background knowledge.
- There are *many* different Bayesian accounts of this form. Earman surveys several of these, and presents his own "clarified" version.
- I will briefly sketch an account due to Ellery Eells (in the course reader, taken from his book *Rational Decision and Causality*).
- Eells makes three assumptions about K (commonly made in here)

$\Pr(Ra \mid H \,\&\, K) = \Pr(Ra \mid K)$	$\Pr({\sim}Ra \mid {\sim}Ba \,\&\, K) \approx 1$	$\Pr(Ba \mid Ra \,\&\, K) \not\approx 1$

"Success Stories" of Bayesian Confirmation IV

- Eells then adopts the *difference measure* of degree of confirmation. That is, he uses $d(H, E \mid K) = \Pr(H \mid E \,\&\, K) - \Pr(H \mid K)$.
- Eells then argues that, given his three assumptions, we have:

$$d(H, E \mid K) = \Pr(H) \cdot \left(\frac{1}{\Pr(Ba \mid Ra \,\&\, K)} - 1 \right)$$
$$> \Pr(H) \cdot \left(\frac{1}{\Pr({\sim}Ra \mid {\sim}Ba \,\&\, K)} - 1 \right) = d(H, E' \mid K)$$

- This is exactly the result we wanted: that a black raven confirms that all ravens are black *more strongly* than a nonblack nonraven does.
- I think Eells' account is one of the simpler accounts out there. But, it seems to me that it is not entirely rigorous. There seem to be gaps in parts of his derivation (Ellery intended this — it was only intended as a "sketch"). **Paper Topic**: make Eells' argument "sketch" fully rigorous.

"Success Stories" of Bayesian Confirmation V

- It is often said that "novel" or "surprising" evidence should provide better support than "unsurprising" or "expected" evidence.
- Bayesianism gives us a way to explain this intuition. On a Bayesian account, if evidence is *improbable*, then it is also *surprising* (if learned).
- So, it would be nice if we could show that (other things being equal) *improbable evidence confirms a hypothesis more strongly than probable evidence does.* Or, more formally and generally, we'd like to show that:

  (*) If $\Pr(E \mid K) < \Pr(E' \mid K)$, then $\mathfrak{c}(H, E \mid K) > \mathfrak{c}(H, E' \mid K)$.

- Earman (page 64) explains that (*) *does* hold, assuming (*i*) that E is *deductive evidence* (*i.e.*, that $H \,\&\, K \vDash E$), and (*ii*) that we use the *difference measure* d as our measure $\mathfrak{c}$ of degree of confirmation.
- Does (*) hold *more generally*, and/or for *other* measures of support $\mathfrak{c}$?