

Old Evidence, Logical Omniscience & Bayesianism

Branden Fitelson

Department of Philosophy
Group in Logic and the Methodology of Science
&
Institute for Cognitive and Brain Sciences
University of California–Berkeley

`branden@fitelson.org`
`http://fitelson.org/`

It must in a certain sense be impossible for us to go wrong in logic.

—Wittgenstein

- Naïvely, “the problem of logical omniscience” (for Bayesian epistemology) is generated *via* the following argument:
- Bayesians assume various things about epistemically rational agents, including the following (and more later):
 - Epistemically rational agents have quantitative degrees of belief (credences) in statements [and credences play an important role in epistemology – more on their role later].
 - Credences of epistemically rational agents can be modeled using *probability functions* [*epistemic states* of rational agents can be represented using *probability models*].
- Probability theory implies that “probability functions assign equal probabilities to logically equivalent statements.”
- ∴ Bayesian epistemology implies that epistemically rational agents must assign equal credences to logically equivalent statements. So, *Bayesianism implies logical omniscience*.
- Does this mean that Bayesians can’t model logically *non-omniscient* agents? No. Today, I’ll try to explain why.

- We will need some technical background about the probability calculus. But, I'll try to keep it to a minimum.
- A *probability model* \mathcal{M} consists of a finite (*i.e.*, finitely many atomic sentences) sentential language P , together with a function $\text{Pr} : P \rightarrow \mathbb{R}$ such that for all sentences p and q in P :
 - $\text{Pr}(p) \geq 0$.
 - If $p \models_P \top$, then $\text{Pr}(p) = 1$.
 - If $p \& q \models_P \perp$, then $\text{Pr}(p \vee q) = \text{Pr}(p) + \text{Pr}(q)$.
 - $\text{Pr}(p \mid q) \stackrel{\text{def}}{=} \frac{\text{Pr}(p \& q)}{\text{Pr}(q)}$, provided that $\text{Pr}(q) \neq 0$.
- We will see some salient examples of \mathcal{M} 's in a few slides.
- **Fact.** If $p \models_P q$, then $\text{Pr}(p) = \text{Pr}(q)$. *This is the precise theoretical sense in which “all probability functions assign equal probability to logically equivalent statements.”*
- **Definition.** E confirms H relative to (or in) a probability model \mathcal{M} (where $E, H \in P$) just in case $\text{Pr}(H \mid E) > \text{Pr}(H)$.
- **Fact.** If $\text{Pr}(E) = 1$ (in \mathcal{M}), then E can't confirm *any* H (in \mathcal{M}).
- This fact gives rise to “The Problem of Old Evidence” [7].

- Naïvely, the problem of old evidence is generated *via* three “orthodox” Bayesian *epistemic modeling assumptions*:
 - (1) The epistemic state of a rational agent a at a time t can be faithfully characterized by a probability model \mathcal{M}_t^a .
 - (2) All confirmational judgments a makes at t must *supervene* on \mathcal{M}_t^a . More precisely, a may judge (at t) that E confirms H *only if* E confirms H *relative to* \mathcal{M}_t^a .
 - (3) If a knows that E at t , then (in \mathcal{M}_t^a) $\Pr(E) = 1$.
 Assumptions (1)-(3) + **Fact** lead us to an odd conclusion:
 - (4) If a knows that E at t , then (at t) a may *not* judge that E confirms H (and this holds for *any* H in \mathcal{M}_t^a).
- Most Bayesians (myself included) respond by denying (2) [2].
- Some recommend expanding the supervenience base in (2) to include *historical* ($t' < t$) epistemic Pr-models $\mathcal{M}_{t'}^a$.
- Others advise expanding (2)'s SB to include *counterfactual* (e.g., a' is a *counterpart* of a) epistemic Pr-models $\mathcal{M}_{t'}^{a'}$.
- I think we need to expand (2)'s SB to include *objective* probability models, but *that's* a story for another talk [4]!

- The canonical example of “the problem of old evidence” involves Einstein, GTR, and the perihelion of Mercury.
- Einstein (*a*) knew in 1915 (*t*) – and this had *long* been known [13] – that (*E*) the perihelion of Mercury advances at $\approx 43''$ of arc per century (above and beyond the precession already predicted by Newtonian theory). Thus, in \mathcal{M}_t^a , $\Pr(E) = 1$.
- But, contrary to (4), Einstein does (in 1915) *seem* to judge that *E* confirms *H* (GTR+Auxiliaries), and this *seems* to be a reasonable judgment for Einstein to have made at that time.
- As I said, most Bayesians try to find a way to reject (2) here. I have my own way to reject (2) *via* IL [4]. I won't get into it.
- Garber [6] and Jeffrey [9] *accept* that Einstein should *not* have judged (in 1915) that *E* confirms *H*. They offer a different explanation of Einstein's confirmational judgment.
- Idea: Einstein *did* know *E* at *t*, but he *didn't* know (at *t*) that “*H* entails *E*” (he was not “logically omniscient”). So, while *E* *couldn't* have confirmed *H* (at *t*, for *a*), “*H* entails *E*” *could*.

- Next: Garber's "logical learning" approach to "old evidence". But, first: "logical ignorance" and Bayesian coherence.
- There are (at least) *three grades* of logical ignorance:
 - (LI₁) Ignorance of some logical relations *in P* caused by *a* having a false conception of the nature of logic itself.
 - In our present context, this would involve \models_P being an incorrect theoretical explication of logical equivalence in *P*.
 - (LI₂) Ignorance of some logical relations *external to P*, reflected in "representational impoverishment" of *P*.
 - *P* is given an *extrasystematic interpretation* (involving some richer theory *T*), which obscures some *extrasystematic* entailments (\models_T 's). [No *systematic* \models_P -ignorance here!]
 - (LI₃) Ignorance of some logical relations *in P* caused by error, laziness, computational/intellectual limitations, *etc.*
 - This involves *a* failing (at *t*) to recognize some classical tautological equivalences *in P* (*i.e.*, *systematic* \models_P -ignorance).
- 👉 Of these three grades, **only** (LI₃) *can be a cause of classical Bayesian incoherence* (vulnerability to "Dutch Book" [10])!
- I will focus on (LI₂). Few Bayesians worry about (LI₁) [14].

- Garber rejects the “global reading” of (1). He argues that *various* “local” probability models may be appropriate for modeling *various aspects of* the epistemic state of a at t .
- Garber proposes a class of probability models for the purpose of modeling *certain aspects of* Einstein’s epistemic state in 1915 [including his (LI₂)-ignorance of $H \models_T E$].
- Garber’s models \mathcal{G} involve a language P with four atomic statements: A, B, C, D . Initially, $A-D$ are uninterpreted and so *any* credence function Pr over P is rationally permissible.
- Next, Garber *extrasystematically interprets* A as H (GTRA), B as E (mercury data), C as $H \models_T E$, and D as $H \models_T \sim E$. The basic conjunctions of P then become *candidate epistemic possibilities for a at t* , and Pr encodes a ’s *credences at t* .
- “ p is epistemically possible for a at t ” $\stackrel{\text{def}}{=}$ “it is permissible for a to assign $\text{Pr}(p) > 0$ at t .” In this sense, *both* $H \models_T E$ and $H \not\models_T E$ were “possibilities” for Einstein in 1915.
- Now we’re ready to see what Garber’s models \mathcal{G} look like ...

Preliminaries ○○				Old Evidence ○○	Garber ○○●○○	Good ○	Me? ○○	Jeffrey ○○○	References ○
<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	Pr	<i>H</i>	<i>E</i>	$H \models_T E$	$H \models_T \sim E$	$\text{Pr}_{1915}^{\text{Einstein}}$
⊤	⊤	⊤	⊤	$p_1 \in [0, 1]$	⊤	⊤	⊤	⊤	0
⊤	⊤	⊤	⊥	$p_2 \in [0, 1]$	⊤	⊤	⊤	⊥	$p \in (0, 1)$
⊤	⊤	⊥	⊤	$p_3 \in [0, 1]$	⊤	⊤	⊥	⊤	0
⊤	⊤	⊥	⊥	$p_4 \in [0, 1]$	⊤	⊤	⊥	⊥	$q \in (0, 1)$
⊤	⊥	⊤	⊤	$p_5 \in [0, 1]$	⊤	⊥	⊤	⊤	0
⊤	⊥	⊤	⊥	$p_6 \in [0, 1]$	⊤	⊥	⊤	⊥	0
⊤	⊥	⊥	⊤	$p_7 \in [0, 1]$	⊤	⊥	⊥	⊤	0
⊤	⊥	⊥	⊥	$p_8 \in [0, 1]$	→ ⊤	⊥	⊥	⊥	0
⊥	⊤	⊤	⊤	$p_9 \in [0, 1]$	⊥	⊤	⊤	⊤	$r \in (0, 1)$
⊥	⊤	⊤	⊥	$p_{10} \in [0, 1]$	⊥	⊤	⊤	⊥	$s \in (0, 1)$
⊥	⊤	⊥	⊤	$p_{11} \in [0, 1]$	⊥	⊤	⊥	⊤	$t \in (0, 1)$
⊥	⊤	⊥	⊥	$p_{12} \in [0, 1]$	⊥	⊤	⊥	⊥	$u \in (0, 1)$
⊥	⊥	⊤	⊤	$p_{13} \in [0, 1]$	⊥	⊥	⊤	⊤	0
⊥	⊥	⊤	⊥	$p_{14} \in [0, 1]$	⊥	⊥	⊤	⊥	0
⊥	⊥	⊥	⊤	$p_{15} \in [0, 1]$	⊥	⊥	⊥	⊤	0
⊥	⊥	⊥	⊥	$p_{16} \in [0, 1]$	⊥	⊥	⊥	⊥	0

- Garber's aim was simply to describe models \mathcal{G} in which $H \models_T E$ could confirm H , even though $\Pr(E) = 1$ in \mathcal{G} .
- Garber does not provide us with specific *constraints* on p, \dots, u which would *entail* that $\Pr(H \mid H \models_T E) > \Pr(H)$.
- Jeffrey [9] and Earman [1] pick-up where Garber leaves off, and they each give sufficient conditions in this sense:
 - **Jeffrey:** $\Pr(H \models_T E \vee H \models_T \sim E) = 1$, i.e., $q = 0$ and $u = 0$. Then, $\Pr(H \mid H \models_T E) > \Pr(H)$ reduces to $\frac{p}{p+r+s} > p$, which follows from $p, \dots, u \in (0, 1)$ and $p + q + r + s + t + u = 1$.
 - **Earman:** $\Pr(H \mid H \models_T E) > \Pr(H \mid H \not\models_T E \ \& \ H \not\models_T \sim E)$. Algebraically, this reduces to: $\frac{p}{p+r+s} > \frac{q}{q+u}$. Non-trivially, this entails $\frac{p}{p+r+s} > p + q$ [viz., $\Pr(H \mid H \models_T E) > \Pr(H)$], because $p, \dots, u \in (0, 1)$ and $p + q + r + s + t + u = 1$.
- Earman's constraint is plausible (for Einstein in 1915).
- This gives Garber a plausible story: In 1915, Einstein learned that $H \models_T E$ (this is also plausible [13]), and it was *this* which boosted Einstein's credence in H . E did not provide any boost (in 1915), since he already knew it.

- As any classical Bayesian must, Garber is assuming that Einstein is omniscient in sense (LI₃). That is, he is assuming omniscience about \models_P , where P is the language of G .
 - Garber also assumes Einstein has *a modicum of* (high-level) knowledge about \models_T . This (incomplete!) extrasystematic logical knowledge is reflected in G 's probability function Pr .
 - Garber uses an idealized, “local” probability model over P to model learning logical relations in T . Modeling P -logical learning would (presumably) require *another* “local” model.
-
- Is Garber's “extrasystematic interpretation” of P (inducing “extrasystematic relations” between P 's *atoms* to partially reflect logical relations *external to P*) *kosher*? Well, it had better be!
- 👉 Historically, this is a *central* Bayesian technique ([3],[11]).
-
- Paradox? Q: How can $H \models_T E$ and $H \not\models_T E$ both be *epistemic* possibilities for a when a *knows* they *can't* both be *logical* possibilities? A: Not all “epistemic possibilities” (in our Garberian sense) express logically possible propositions!

- Good [8] suggests an alternative “evolving probability” (EP) approach, which requires only that *known* logical truths get credence 1, and *known* incompatibles satisfy additivity.
- (EP) is *ambiguous* between grades (LI₂) and (LI₃) ignorance. On an (LI₃) reading, EP amounts to weakening $p \models p q$ to “*a* knows that $p \models p q$ at *t*” [i.e., $K(p \models p q)$] in the Pr-axioms.
- [Is $K(p \models p q)$ an *equivalence relation*? If not, then \mathcal{M} isn’t a probability model. See [5] for a nice discussion, and a rigorous, proof-theoretic alternative to Garber and (EP).]
- On *either* reading, EP recommends that we change (*only!*) our credences so as to “reflect the learned logical relation”.
 - *E.g.*, if we learn “*H* entails *E*”, then EP prescribes adopting a new credence function Pr' such that $\text{Pr}'(E | H) = 1$.
- Garber critiques EP. His main complaint: EP can’t handle OE *via* logical learning, since in OE cases we *already* have $\text{Pr}(E | H) = 1$ *before* *a* learns $H \models_T E$. So, in OE cases, EP can’t account for *any* Pr shift arising from learning $H \models_T E$.
- I think Garber is right. But, I want to say a bit more here.

- Kukla [12] defends EP against Garber. But, he concedes:
On the EP account, ... prior to [learning the logical fact in question], our probability function was incoherent in the classical Bayesian sense. Indeed, our probability functions are *always* incoherent ... and no doubt always will be.
- This is a brave concession! And, a mistake. **Incoherence** requires (LI₃)-ignorance. Making this grade of ignorance so ubiquitous just *trivializes* the notion of “coherence”.
- *Pace Good and Garber* (& Jeffrey), I suspect we can't always repair (capture) all effects (aspects) of logical ignorance (learning) at *t merely* by changing our Pr over our old *P*.
- Some examples of logical learning seem to involve moving to *new language P'*, which can articulate logical relations obscured in *P*. This goes beyond previous approaches.
- Garber models varying degrees of logical ignorance about *T* by tweaking his extrasystematic interpretation of *P*. This is clever, and an improvement over EP. But, I think this still obscures a salient kind/facet of change in epistemic state.

- OK, so the question is: if we sometimes want not only a Pr-shift but also a P -shift (i.e., a *model* shift), then what principles should guide us in formulating our new model?
- This is not an easy question. But, I have a few ideas.
- Like Garber, I suggest adding atoms to the (naïve) model (language) so as to capture obscured logical structure.
- But, I suggest: (a) do this *diachronically*, and (b) reflect the learned relations as *tautological relations in P' itself* (vs “ \models_T -relations” in an *extrasystematic interpretation of P*).
- *E.g.*, when a learns $H \models_T E$, we might model a as moving to a new $\mathcal{M}' [\langle P', \text{Pr}' \rangle]$ in which what was expressed in P as “ H ” (GTRA) is now expressed as “ $E \& X$ ”, for a new X in P' .
- If this is an OE case, then $\text{Pr}(E)$ and $\text{Pr}'(E)$ will both equal 1. Thus, $\text{Pr}'(E \& X) = \text{Pr}'(X)$. So, the “probability boost” Garber wants reduces to $\text{Pr}'(X) > \text{Pr}(H)$. *I.e.*, in old-evidence cases:
 - 👉 Learning $H \models_T E$ between t and t' boosts a 's credence in H if the part of H that “goes beyond” E (represented by a new “ X ” in P') has greater credence for a at t' than H had at t .

- Garber introduces a new proposition $[C]$ into his algebra, which “expresses” $H \vDash_T E$. This allows Garber to model the (logical) learning of $H \vDash_T E$ via *conditionalization on C*.
- Jeffrey rejects conditionalization (as a rule) for various reasons. (i) We shouldn’t (always) assign probability 1 to learned contingents. (ii) There isn’t always a proposition in one’s algebra which expresses that which one has learned.
- In OE cases of *logical learning* (i) is moot (here, all is learned *with certainty*). (ii) doesn’t apply to the (empirical) learning of E . Perhaps (ii) applies to the (logical) learning of $H \vDash_T E$.
- But, why *couldn’t* there be a statement in P “expressing” $H \vDash_T E$? No wff in P *systematically* expresses $H \vDash_T E$. So?
- 👉 *That **can’t** be required* of a conditionalization approach ([3], [11]). In any case, Jeffrey abandons conditionalization here.
- One more issue: Like all the others, Jeffrey models logical learning events as updates of Pr ’s over *fixed languages*.
- Putting these worries aside, here’s Jeffrey’s approach ...

- Jeffrey [9, Postscript] operates with a more parsimonious language P_J , containing just two atomic statements: H, E .
- He models the learning of $H \models_T E$ (logical) and E (empirical) as (Jeffrey!) updates on Pr's over P_J : $\text{Pr}^0 \mapsto \text{Pr}^1 \mapsto \text{Pr}^2$.
- We begin with what Jeffrey calls the “ur-function” Pr^0 , which assigns non-extreme credence to each basic conj. of P_J .
- Provisionally, Jeffrey has a do their logical update “first”. He assumes two things about this $\text{Pr}^0 \mapsto \text{Pr}^1$ logical update:
 - (9) $\text{Pr}^1(H) = \text{Pr}^0(H)$. *Why?* “Learning that H implies something that may well be false neither confirms nor infirms H ”.
 - (10) $\text{Pr}^1(H \& \sim E) = 0$. *Why?* Because a learned $H \models_T E$ here!
- Jeffrey-updating subject to (9)&(10) yields a *unique* Pr^1 from Pr^0 :

H	E	Pr^0		H	E	Pr^1
⊤	⊤	$a \in (0, 1)$	$H \models_T E$ \mapsto	⊤	⊤	$a + b \in (0, 1)$
⊤	⊥	$b \in (0, 1)$		⊤	⊥	0
⊥	⊤	$c \in (0, 1)$		⊥	⊤	$c \in (0, 1)$
⊥	⊥	$d \in (0, 1)$		⊥	⊥	$d \in (0, 1)$

- “Next”, the empirical update (E) occurs. Jeffrey models this as learning E *with certainty*. [Yep, by **conditionalizing** Pr^1 on E ! He must do so *otherwise the OEP does not arise*. Not very “radical”!]
- In all, Jeffrey makes 3 more assumptions about the 2 updates:
 - (11) $\text{Pr}^2(\cdot) = \text{Pr}^1(\cdot | E)$. *Why?* E is learned *with certainty* in OE.
 - (12) $\text{Pr}^0(H | E) = \text{Pr}^0(H)$. *Why?* Unclear (simplifies the math).
 - (13) The logical and empirical updates should *commute* — the order in which they come should not have an effect on Pr^2 . *Why?* Otherwise, E cannot be “old” when $H \models_T E$ is learned.
- Jeffrey-updating subject to (9)–(13) yields a *unique* Pr^2 from Pr^0 :

H	E	Pr^2
\top	\top	$\frac{a+b}{a+b+c} \in (0, 1)$
\top	\perp	0
\perp	\top	$\frac{c}{a+b+c} \in (0, 1)$
\perp	\perp	0

- Since $\frac{a+b}{a+b+c} > a + b$, $\text{Pr}^2(H) > \text{Pr}^0(H)$, and H has received a “probability boost” from learning (E and then) $H \models_T E$.

- [1] J. Earman, 1992, *Bayes or Bust?*, MIT Press.
- [2] E. Eells, 1985, "Problems of Old Evidence", *Pacific Phil. Quarterly*, **66**: 283-302.
- [3] B. Fitelson and J. Hawthorne, 2005, "How Bayesian Confirmation Theory Handles the Paradox of the Ravens", forthcoming in *Probability in Science*, E. Eells and J. Fetzer, eds., Open Court. [fitelson.org/ravens.pdf]
- [4] B. Fitelson, 2005, "Logical Foundations of Evidential Support", forthcoming in *Philosophy of Science*. [fitelson.org/psa2004.pdf]
- [5] H. Gaifman, 2004, "Reasoning with Bounded Resources and Assigning Probabilities to Arithmetical Statements", *Synthese*, **140**: 97-119.
- [6] D. Garber, 1983, "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory", in *Testing Scientific Theories*, J. Earman, ed., 99-132.
- [7] C. Glymour, 1980, *Theory and Evidence*, Princeton University Press.
- [8] I.J. Good, 1968, "Corroboration, Explanation, Evolving Probability, Simplicity and a Sharpened Razor", *British Journal for the Phil. of Sci.*, **19**: 123-143.
- [9] R. Jeffrey, 1992, "Bayesianism with a Human Face", in his *Probability and the Art of Judgment*, Cambridge University Press, 77-107.
- [10] J. Kemeny, 1955, "Fair Bets and Inductive Probabilities", *JSL*, **20**: 263-273.
- [11] J. Kemeny, et al., 1957, *Introduction to Finite Mathematics*. Prentice-Hall.
- [12] A. Kukla, 1990, "Evolving Probability", *Philosophical Studies*, **59**: 213-224.
- [13] N. Roseveare, 1982, *Mercury's Perihelion from Le Verrier to Einstein*, OUP.
- [14] B. Weatherson, 2003, "From Classical to Constructive Probability", *NDJFL*.