

Hindsight Bias and Standard Bayesianism: Chapters 7 and 4

Kevin Dorst (kmdorst@mit.edu)

24.805, Fall 2025

I. Ambiguous Bayesianism

Formalization of HB: $\mathbb{E}_{P_a}(P(q)) < \mathbb{E}_{P_a}(P(q)|q)$

Example: balanced-spoon case. How likely to *fall*?

If your prior probability was your sampling propensity for *fall*,¹ you can be unsure about that.

Since your intuitive-physics engine is *good*, you should trust your judgment about this. So exhibit HB.

Fact. P_a exhibits HB given q iff P_a thinks $P(q)$ and q are correlated.

'Correlated' here means *across the possibilities you leave open*, not generically across propositions and times.

[Draw 3-class $N-n$ frame.]

Worry: Does HB make your estimate for your credence less accurate? Not necessarily—since you might've *mis-estimated* your prior.²

General predictions: the two primary drivers of HB will be:
(1) ambiguity and (2) self-trust.

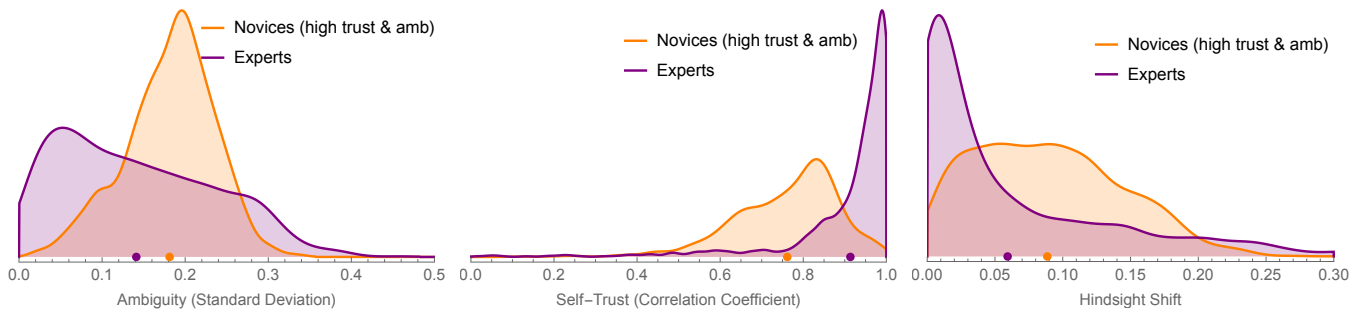
II. Explaining existing findings

Expertise. Expertise will *decrease ambiguity* but *increase self-trust*. Depending on which effect predominates, can decrease or increase HB.

→ When novices have hi amb and trust expertise will decrease HB.

→ When novices have lo amb and trust, expertise will increase HB.

Simulations:



¹ Or some other hard-to-discern disposition that uses your intuitive-physics engine

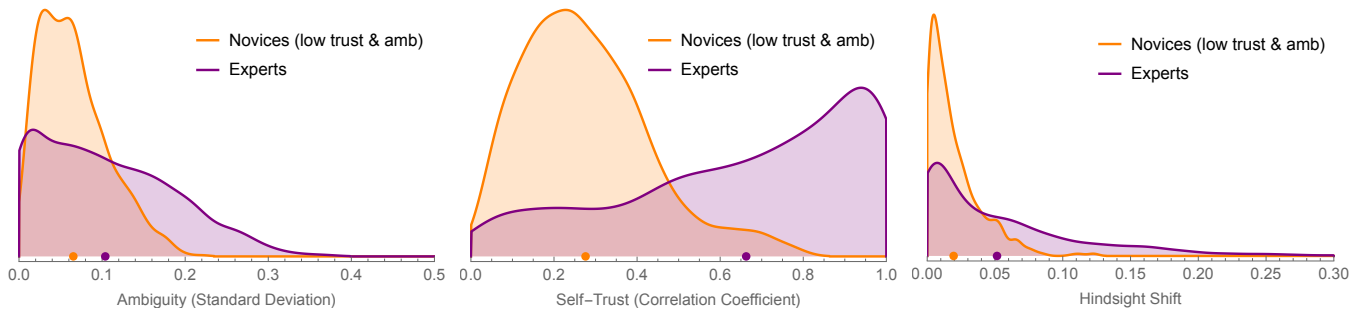
$\mathbb{E}_{P_a}(P(q)) < \mathbb{E}_{P_a}(P(q)|q)$ iff $\text{cov}_{P_a}(P(q), q) > 0$.

In that sense, SBs never think their opinions are correlated with truth.

² If the prior is 'valuable' (Ch. 6), it increases accuracy in expectation. But it's inevitable that *sometimes* HB will make you less accurate about what your credence was. And if do for many separate questions, maybe predictably so (Ch. 8).

social prediction?

chess novices



Trivia. Trivia questions tend to be ones where you have uncertainty about *both* (1) what evidence you would've had (called to mind), and (2) how to interpret it. Case histories and forecasts usually only (2).

III. New Predictions

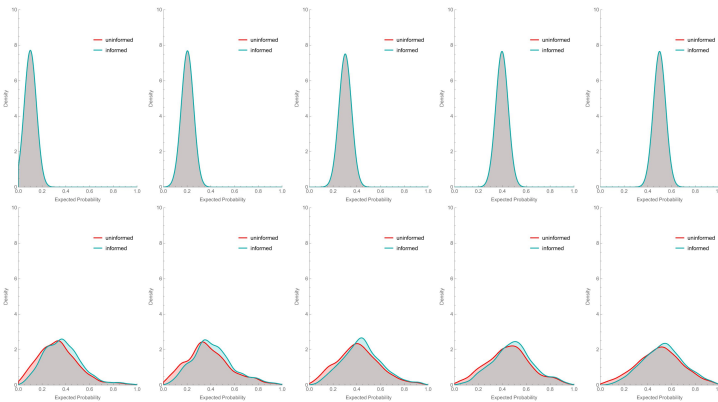
Replication crisis. Want better (1) theory and (2) experiments.

(Pre-registration; 'random-effects' models)

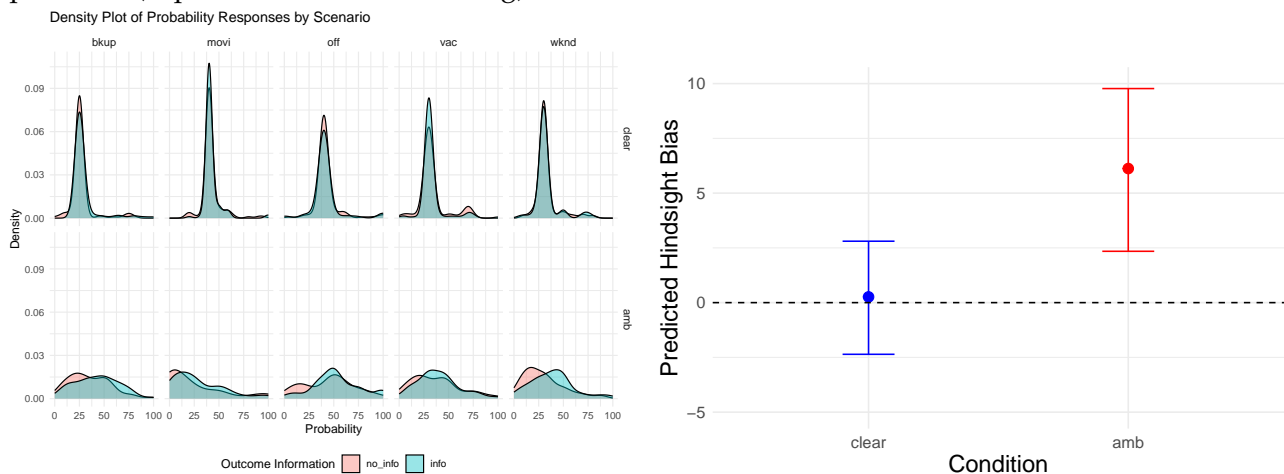
Experiment 1: Removing ambiguity should remove hindsight bias

Procedure, including comprehension checks. Vignette example.

Simulations:



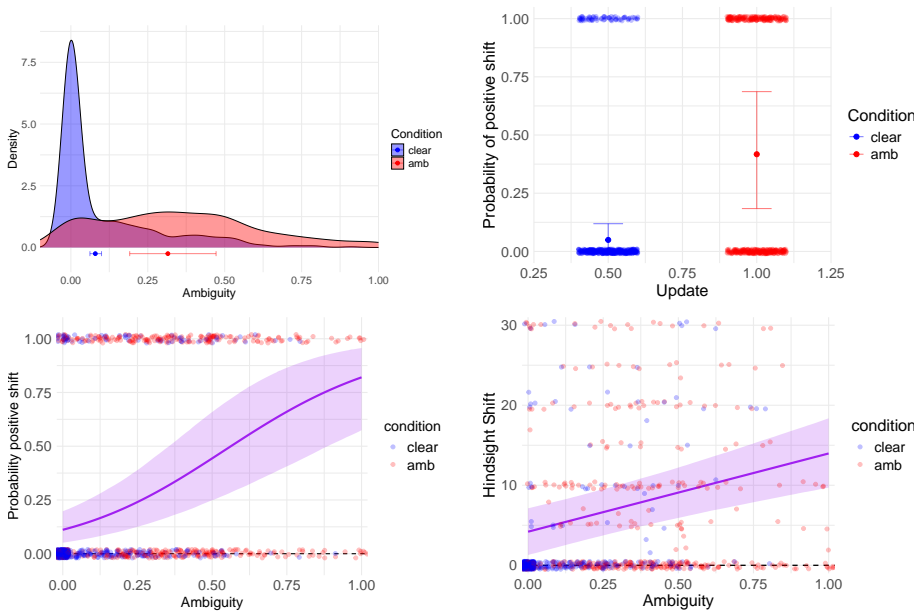
Experiment (explain what model is doing):



Experiment 2: Within-subject, removing ambiguity

Train to distinguish guess from ‘what they really think’; comp checks.

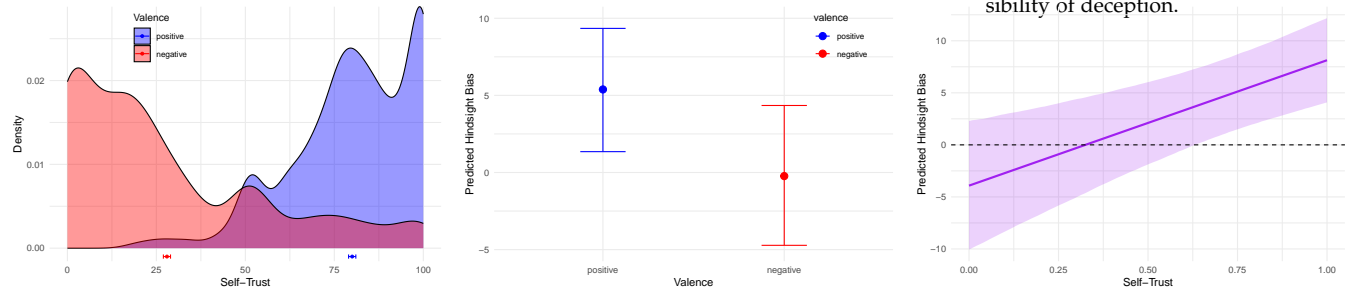
Lose more people—so worry about selection effects.



Experiment 3: reducing self-trust should reduce/eliminate HB

City-size comparisons; use deception to manipulate self-trust.

Note: consent form must mention possibility of deception.



New ideas: Moral luck?

IV. Standard Bayesianism

Self-Awareness. (‘Non-Delusion?’) P is *self-aware* iff never rules out having its actual opinions.

Standard Bayesianism: A frame (W, α, P, P^+) is *Standard-Bayesian* iff:

- i) P is self-aware and clear;³ and
- ii) There is a partition \mathcal{Q} such that your posterior is the result of conditioning on the true cell of \mathcal{Q} : for all w and q , $P_w^+(q) = P_w(q|\mathcal{Q}_w)$.

Why partitional updates?

Claim: Under clarity, minimally-rational updating is partitional updating. Why?

‘What posterior you ended up with’ always forms a partition; you’re always certain of that; and that always captures what you learned.

For all w , $P_w(P = P_w) > 0$.
 Guaranteed but factivity: $P_w(w) > 0$,
 i.e. $\langle P(q) = 1 \rangle \rightarrow q$. But weaker.

³ Equivalently: P is always correctly certain about itself. (Self-awareness rules out P being certain but mistaken.)

Big debate brewing in formal epistemology...

Exogenous vs. total evidence. Flip coin; asymmetric reporting.
Island of trials (Monty Hall).

Generalizing:

No Foregone Conclusions: If you should be certain that your posterior estimate for X will be high, you should already have a high estimate for X .

Formally: if $P_w(\mathbb{E}_{P^+}(X) \geq x) = 1$, then $\mathbb{E}_{P_w}(X) \geq x$.

Clarity + No Foregone Conclusions \Rightarrow partitional (so SB) updating.

Without clarity, things get more contested.

E.g. unmarked clock. Know that you'll be unsure of the strongest thing you're certain of.

V. Martingale and Reflection

We know Standard Bayesians validate Martingale: expectations for your future credence have to equal your current credence.

In fact, they satisfy a stronger principle:

Reflection: Completely defer to your future self.

Conditional on your posterior having credence x in q and y in p , have (conditional) credence x in q .

Formally: $P_w(q | \langle P^+(q) = x \rangle \& \langle P^+(p) = y \rangle) = x$.

Standard-Bayesianism validates Reflection (Briggs 2009; Weisberg 2007).

In fact, the reverse is true. If we assume Reflection⁴, then SB follows.

Some people think this is an argument *for* Standard Bayesianism.

I think it's an argument *against* Reflection, since it shows that Reflection only makes sense under clarity.

Suppose P^+ is ambiguous, so you leave open a w in which $P_w^+(p) = 0.7$ but $P_w^+(P^+(p) = 0.7) = 0.9$.

A special case of Reflection: $P_a(P^+(p) = 0.7 | P^+(P^+(p) = 0.7) = 0.9) = 0.9$

But Reflection says it doesn't matter if we tack on another conjunct:

$P_a(P^+(p) = 0.7 | \langle P^+(P^+(p) = 0.7) = 0.9 \rangle \& \langle P^+(p) = 0.7 \rangle) = 0.9$

References

- Briggs, Ray, 2009. 'The Anatomy of the Big Bad Bug'. *Nous*, 43(3):428–449.
Gaifman, Haim, 1988. 'A Theory of Higher Order Probabilities'. In Brian Skyrms and William L Harper, eds., *Causation, Chance, and Credence*, volume 1, 191–219. Kluwer.
Isaacs, Yoaav and Russell, Jeffrey Sanford, 2022. 'Updating Without Evidence'. *Nous*, (March):1–33.
Samet, Dov, 1999. 'Bayesianism without learning'. *Research in Economics*, 53:227–242.
Schoenfield, Miriam, 2017. 'Conditionalization Does Not (In General) Maximize Expected Accuracy'. *Mind*, 126(504):1155–1187.
Weisberg, Jonathan, 2007. 'Conditionalization, reflection, and self-knowledge'. *Philosophical Studies*, 135(2):179–197.

Schoenfield 2017

Draw naive and sophisticated updates.

Problem: Ambiguous-Bayes is going to violate this. :/

Possibly: Talk through how Schoenfield algorithm would work here. See Isaacs and Russell 2022.

Special case: let p be a tautology and $y = 1$, so second conjunct drops out:
 $P_w(q | P(q) = x) = x$

⁴ And also assume it applies synchronically: P reflects P itself

E.g. Gaifman 1988; Samet 1999.

Letting $r = \langle P^+(p) = 0.7 \rangle$, this says
 $P_a(r | P^+(r) = 0.9) = 0.9$