

Hindsight Bias: Hedden 2019 and Chapter 7

Kevin Dorst (kmdorst@mit.edu)

24.805, Fall 2025

I. The Findings

'Knew it all along' effect.

Matters for retrospective assessments of self and others.

Within-subject designs

Between-subjects (informed vs. uninformed subjects) designs.

Between-subjects designs are direct evidence for:

Immediate hindsight bias: When you learn q , updating from P_a to P_a^+ , that raises your expectation for your prior credence¹ in q , $P(q)$.

$$\mathbb{E}_{P_a}(P(q)) < \mathbb{E}_{P_a^+}(P(q))$$

Within-subjects designs are (usually) evidence for:

Delated hindsight bias: When you learn q , updating from P_a to P_a^+ , that raises your expectation for your *earlier* credence in q , $P^-(q)$.

$$\mathbb{E}_{P_a}(P^-(q)) < \mathbb{E}_{P_a^+}(P^-(q))$$

Empirical trends (from meta-analyses):

Relevance. Personal relevance/valence don't increase hindsight bias.

Expertise. Expertise is not well-correlated with hindsight bias.

Trivia. People exhibit greater hindsight bias for trivia questions.

Resilience. There are few reliable methods² to reduce hindsight bias.

Eg Clinton impeachment before and after

Eg Tetlock2005;
Fischhoff on trivia questions;
doctors w/ case history rating Hep.

¹ Or: the rational prior credence

² Other than 'consider the opposite', which is more general. (Does Gerken contest Resilience? What's he say?)

II. Existing Theories

Standard psychological theories are broadly mechanistic:

Motivated reasoning. People want to make themselves look good.

→ Has trouble with Relevance, Trivia, and Expertise.

Selective-recall and sense-making. Knowing q strengthens associative connections to q -supporting evidence and memories.

→ Selective-recall has trouble with *immediate* HB.

→ Trivia (random facts) is not subject-matter for sensemaking, so not clear it can explain the bigger effect there.

Metacognitive Fluency. Monitoring yourself for fluency; feel more fluency when you know q so get more confident³ in your judgment.

→ Has trouble with Trivia, Expertise, and clarity reducing HB.

³ Extreme?

Maybe fluency-differential between informed and uninformed conditions?

Hedden: 'Um. Isn't this just Bayesian updating?'

- If Bayesian, updating is conditioning: $P_a^+ = P_a(\cdot|q)$.
- Focus on case where \mathcal{P} is the rational credence⁴ for some agent A —say, your prior self or person being evaluated.
- It's reasonable to be unsure what \mathcal{P} is: you might be unsure of both (1) what evidence A has and (2) what body of evidence e supports.
 - (1): prior self in within-subject designs. Railroad company in negligence case.
 - (2): complex formula (tautology or contradiction?); complicated evidence (regression support strong or weak effect?); balancing theoretical virtues.
- Usually should take $\mathcal{P}(q)$ to be correlated with truth-value of q .
 - (1): Don't expect random bodies of evidence to be misleading.
 - (2): These a priori facts are *from your subjective perspective* correlated with the truth—if the best regression says chemical likely causes cancer, you should raise your credence that it causes cancer.
- So, usually, if $P_a(q) = 0.6$, then $P_a(q|\mathcal{P}(q) = 0.8) > 0.6$.
- Correlations are symmetric.
 - If $P_a(q|\mathcal{P}(q) = 0.8) > P_a(q)$, then $P_a(\mathcal{P}(q) = 0.8|q) > P_a(\mathcal{P}(q) = 0.8)$.
 - Likewise for other x^+ higher than $P_a(q)$, so your credence that $\mathcal{P}(q)$ was these higher numbers should go up.
 - ⇒ Learning q raises estimate of $\mathcal{P}(q)$.

Worries:

Gerken(?): \approx

- (1) ψ ists are broadly aware of this, and try to control for it⁵, and
- (2) the fact that HB is reduced by (the right sort of) thinking/reflection⁶ suggests that it's not ideally rational.

Kevin:

- Generally: people *are* aware that something that *looks* like HB—but with more-idealized probability functions (like the objective chances)—can be rational. They think there's something bad about it being done on (functions close to) *your prior*.
- **Q:** is \mathcal{P} 'close enough' to P to capture the phenomenon?
 - Want an X that:
 - (i) you're reasonably uncertain about,
 - (ii) is correlated with the truth, and
 - (iii) captures the range of findings.
- *Rational credence function* satisfies (i) and (ii) but struggles with (iii).
 - When arguing for (i) and (ii), natural to idealize \mathcal{P} —talking about the ideal version of yourself which has done all the calculations right.
 - But when arguing for (iii), important not to idealize too much.
- If switch to your *actual* (prior) credence function, then we need higher-order uncertainty (P unsure of P), rather than probabilistic uncertainty (P unsure of \mathcal{P} ⁷).

⁴Or evidential support function.

Brian uses cr for your credence function and ES_A for the evidential support function for the target agent A .

If \mathcal{P} is your *informed* probability \hat{P} , you'll Reflect it: $P_a(q|\mathcal{P}(q) = x) = x$. So this will hold.

Analogy: bias is evidence of landing heads; so heads is evidence of bias.

⁵ Eg by clearly specifying the evidence, or having it be the same person, and by giving you *lots* of evidence about what evidence A had. [But sometimes more evidence reduces clarity!]

⁶ Namely, 'consider-the-opposite' debiasing

eg for between-subjects cases (or blame/negligence cases) want to use what *you* (not your highly-idealized self) would've thought.

⁷ In Brian's notation, cr unsure of ES_A .

→ **Fact.** If use P , Standard Bayesians never exhibit (immediate or delayed) hindsight bias.

But once go HOU, need to state the correlation assumption (i.e. (ii)) carefully to avoid inducing clarity (falsifying (i)).

III. Ambiguous Bayesianism

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

If your prior probability was your sampling propensity for *fall*,⁸ 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. In that sense, SBs never think their opinions are correlated with truth.

[Possibly: draw 3-class N-n frame.]

Worry: Does HB make your estimate for your credence less accurate? Not on average—since you might've *mis-estimated* your prior.⁹

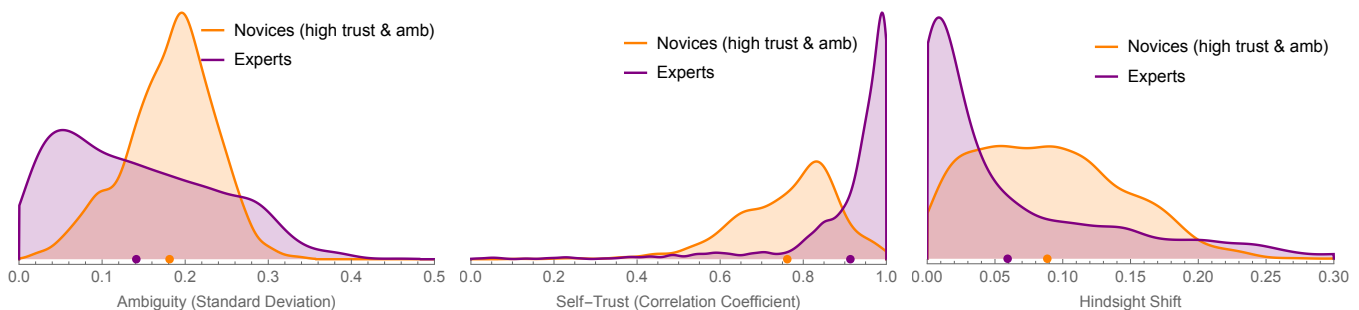
General predictions: the two primary drivers of HB will be (1) ambiguity and (2) self-trust.

IV. 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 ambiguity and trust (social prediction?), expertise will decrease HB.

When novices have lo ambiguity and trust (chess novices), expertise will increase HB.



Can't use $P(q|P(q) = x) = x$.

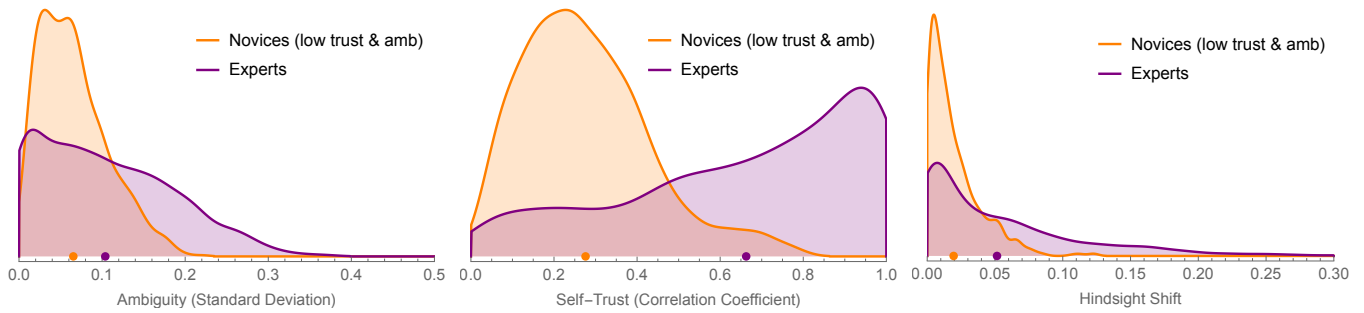
Even Hedden's assumption [if $P(q) = x$ and $x^+ > x$, then $P(q|P(q) = x^+) > P(q)$] arguably too strong. Maybe fine to assume *usually* holds?

This name, 'Ambiguous Bayesianism', is ugly. Is there a better one? Needs to contrast with 'Standard Bayesianism'...

⁸ 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 $\text{cov}_{P_a}(P(q), q) > 0$.

⁹ If the prior is valuable, it always increases accuracy in expectation. But under ambiguity it's inevitable that *sometimes* HB will make you less accurate about what your credence was.



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).

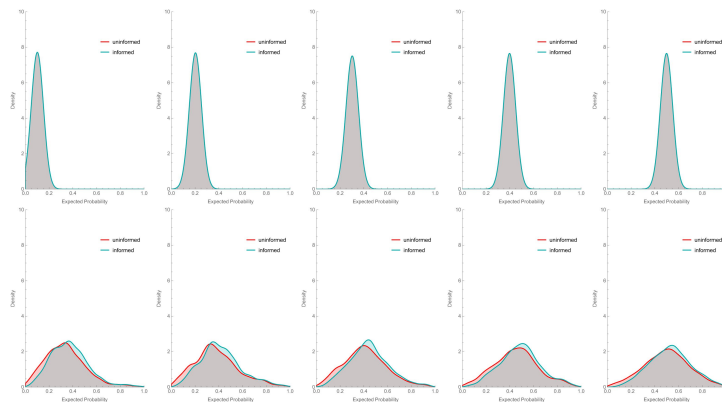
V. Studies [if we get this far...]

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

(Pre-registration; random-effects models)

Experiment 1: Removing ambiguity should remove hindsight bias

Procedure. Vignette example. Simulations:



Experiment:

Density Plot of Probability Responses by Scenario

