Modeling Ambiguity and Probability-Weighting: Chapter 3 and Enke and Graeber 2023

Kevin Dorst (kmdorst@mit.edu) 24.805, Fall 2025

I. Standard Bayesianism

Outcome spaces; propositions; probability distributions as vectors.

Laundry + Thomas

Constants vs. (random) variables.
Random variables used to specify propositions.

2 and x vs. X

Multiple names for a given proposition

Questions/evidence as partitions; conditioning as wiping mud.

Posterior probability P^+ as a *variable probability function*.
Stochastic matrices; use P^+ to specify propositions.

Conditional probability vs. posterior probability.

Prior probability P also as a *variable* probability function.

Not a constant distribution π

Probability frames $(W, \mathfrak{a}, P, P^+)$.
E.g. Thomas example.

$P_\mathfrak{a}$ vs P

Standard Bayesianism (SB) $\approx$ P is constant and there is a partition $\mathcal{Q}$ such that P^+ is the result of conditioning P on the true answer to $\mathcal{Q}$.

For all w, $P^+_w(\cdot) = P_w(\cdot|\mathcal{Q}_w)$, where $\mathcal{Q}_w$ is the true member of $\mathcal{Q}$ at w.

These are the models that are used to prove the convergence theorems.

Fact 1. Standard-Bayesian frames are always *clear*:
If $P_w(q) = x$, then $P_w(P(q) = x) = 1$; and
If $P^+_w(q) = x$, then $P^+_w(P^+(q) = x) = 1$.

Also true of *conditional* probabilities. $P_\mathfrak{a}(r|l) = \frac{2}{3}$, and $P_\mathfrak{a}(P(r|l) = \frac{2}{3}) = 1$.

Why? Because you (1) are sure of what your prior was,[1]
(2) are sure of what you learned,[2] and
(3) are sure that you updated on what you learned by conditioning.[3]

[1] Since P constant

[2] Since $\mathcal{Q}$ is partitional

[3] Since that's true at all worlds

Fact 2. Standard-Bayesian updates validate the 'Martingale' principle.

- Expectations of a random variable X, $\mathbb{E}_{P_\mathfrak{a}}(X)$.
- Expectations used to guide decisions.
- Prior $P_\mathfrak{a}$ can form expectation for posterior P^+, written $\mathbb{E}_{P_\mathfrak{a}}(P^+(l))$
- **Martingale:** Your prior expectation for your posterior probability in q must equal your prior probability for q.
- Denying Martingale leads to an important form of confirmation bias and (thereby) convergence failures.

Not what you 'expect'. What $P_\mathfrak{a}$ is confident average would be in long run.
B^1 = \$1 if did laundry;
B^2 = \$2 if did and Thomas remembers

$\mathbb{E}_{P_\mathfrak{a}}(P^+(q)) = P_\mathfrak{a}(q)$
'evidence of evidence is evidence'

$\mathbb{E}_{P_\mathfrak{a}}(P^+(q)) > P_\mathfrak{a}(q)$ = investigate in a way that you expect, on average, to increase your credence in q.

II. Ambiguous Bayesianism

P is a variable. So let it vary across the worlds it leave open.

$P_w(u) > 0$ and $P_w(v) > 0$, but $P_u(q) \neq P_v(q)$.

Unrealistic that your conditional probabilities about Thomas's memory

were clear: even if $P_{\mathfrak{a}}(r|l) = \frac{2}{3}$, you're not sure of that.
Suppose we want to let $P(r|l)$ vary. How?

Make sure worlds are specific enough to tell us what $P(r|l)$ is.
Draw frame that lets $P_w(r|l)$ vary across worlds.

Genuine higher-order uncertainty.
Unsure whether $\langle P(r) = 0.33\rangle$ or $\langle P(r) = 0.17\rangle$.
Unsure whether $\langle P(r|l) = \frac{2}{3}\rangle$ or $\langle P(r|l) = \frac{1}{3}\rangle$.

In fact, $P_{\mathfrak{a}}(r) = 0.33$.
In fact, $P_{\mathfrak{a}}(r|l) = \frac{2}{3}$.

Informed probabilities, $\widehat{P}$ vs 'uninformed' (subjective) probabilities P.

Factorability: P_w always a weighted average of the possible values of $\widehat{P}$.

Updating: lay down a partition $\mathcal{Q}$ over P, condition on true cell.
(Imagine adding a line to a computer program: 'resample if $\neg\mathcal{Q}_w$'.)

If time: see glimmers of hindsight bias and confirmation bias.

Analogy(?): unsure if (1) angry at work or (2) angry at Thomas. If learned (2), you'd get *more* angry at Thomas. *Better(?):* P dispositional. So learning about P is learning about your dispositions, affecting your dispositions. → You're unsure whether you're able to (disposed to) perform a hand-stand. If you learned that you are, you'd be *more* able to (more likely to succeed).

III. Probability-Weighting

Risk aversion. Standard: diminishing marginal utility.

But that's *not* how real people behave. Four-fold pattern:

Unlikely gains: With low probabilities of gains, people are *risk seeking*.
Unlikely losses: With low probabilities of losses, people are *risk-averse*.
Likely gains: With high probabilities of gains, people are *risk-averse*.
Likely losses: With high probabilities of losses, people are *risk-seeking*.

Buying lottery tickets.

Buying fire insurance.

Preferring (i) a sure \$90 to (ii) a bet that's 95%-likely to yield \$100 and 5% likely to yield \$0.

Preferring (i) a 95%-chance to lose \$100 and 5%-chance to lose nothing to (ii) a sure loss of \$90.

Classic explanations: *prospect theory* (Tversky and Kahneman 1992), or *rank-dependent utility theory* (Quiggin 1982; Buchak 2013).

The idea: add *decision-weights* to modify probabilities.
Philosophers (like Buchak) often focus on simple ones like x^2; but real people show an *inverse-S curve*. That explains the four-fold pattern.

Regressive (flat) in middle; sensitive at endpoints

Enke and Graeber: evidence that distortion in people's elicited subjective probabilities.

Enke and Graeber: explain with *cognitive uncertainty* (CU), i.e. subjective uncertainty over your 'ex ante utility-maximizing decision'.

= expectedly-best option, i.e. rational option to take

Measuring CU: certainty equivalents; estimates of optimal guess.

Model of CU:

- Draw n binomial samples from ideal probability $\mathcal{P}(q)$.
- Default (beta) prior δ, with $\delta(q) = \mathbb{E}_\delta(\mathcal{P}(q))$. δ updated on samples to get elicitation, $\delta(q|samples = \frac{m}{n})$.

Predicts noise and regressive-ness. (Need to vary n at endpoints to get sensitivity at endpoints.)

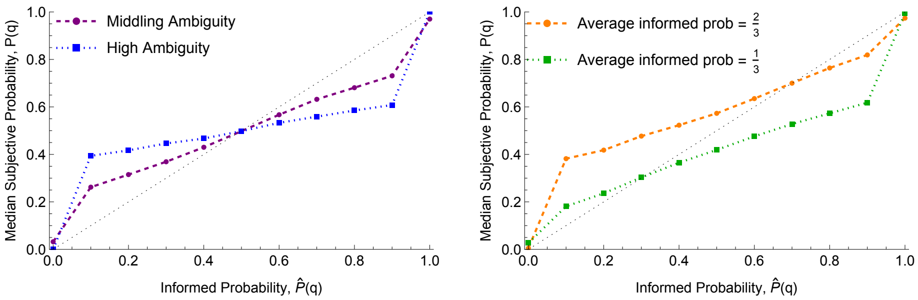
IV. Cognitive Uncertainty as Higher-Order Uncertainty

Basic idea: interpret measures of cognitive uncertainty as measuring *higher-order* uncertainty.

- Generate random (factorable) frames and updates;
- Treat (posterior) informed probability $\widehat{P}_{\mathfrak{a}}(q)$ as objective probability;
- Track (posterior) subjective probability $P_{\mathfrak{a}}(q)$.
- Track (some measure of) higher-order uncertainty.

Simple Ambiguity = 1 - c, where: if $P_{\mathfrak{a}}(q) = x$, then $c := P_{\mathfrak{a}}(P(q) = x \pm 0.01)$.

Bin cases by $\widehat{P}(q)$ $(0, 0.1, 0.2, ...)$, and plot results:



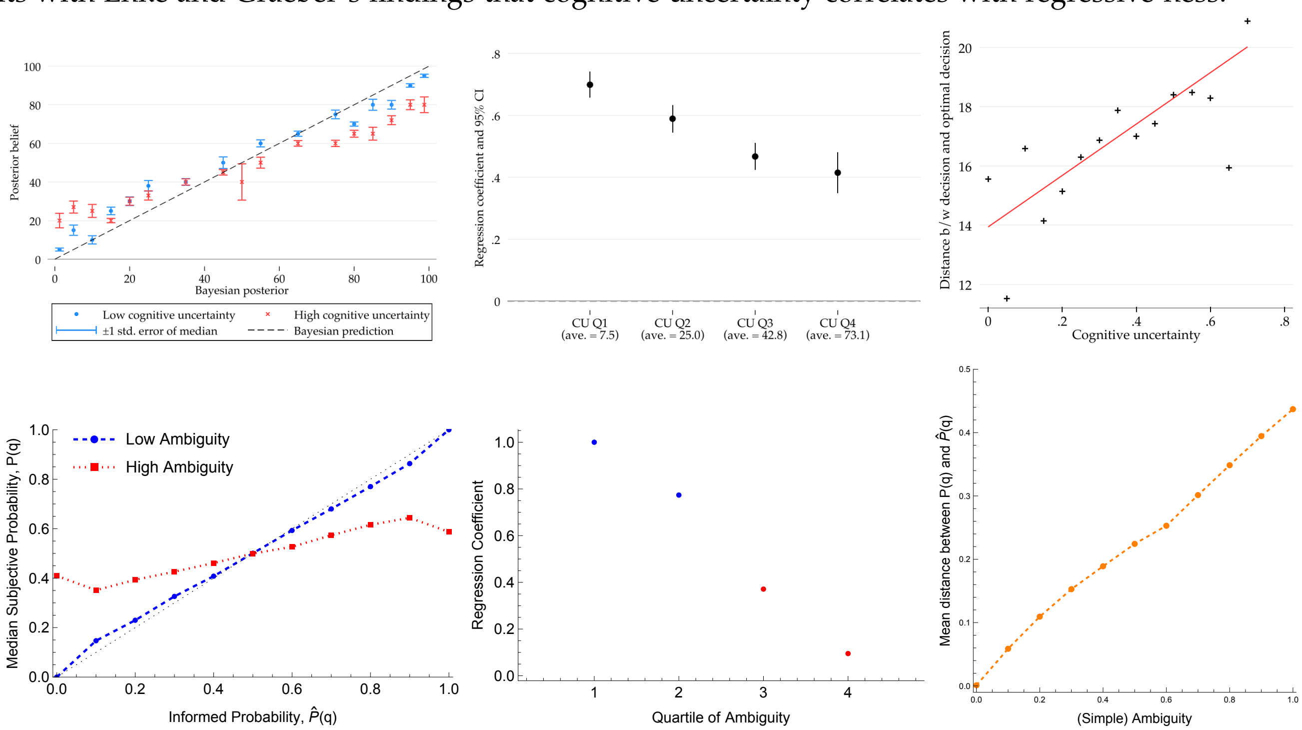
Why?

- *Regressive* because P is a weighted average of both *true* informed probability $\widehat{P}_{\mathfrak{a}}(q)$ and other $\widehat{P}_w(q)$ for other possible w.
  $\rightarrow$ Whether regressive toward 0.5, or higher or lower, depends on what the average informed probability is.
- *Sensitive at end-points* b/c most-common way to get extreme (0 or 1) informed credence in q is to learn something that *settles* whether q.

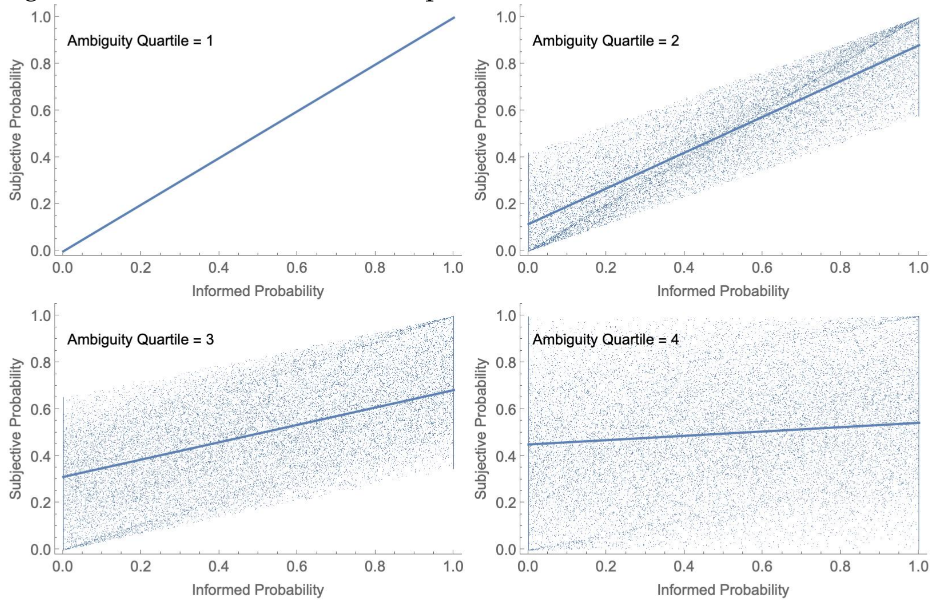
Higher-order doubts skew your probabilities toward the average informed probability

I.e. either $\mathcal{Q}_{\mathfrak{a}} \subset q$ or $\mathcal{Q}_{\mathfrak{a}} \subseteq \neg q$

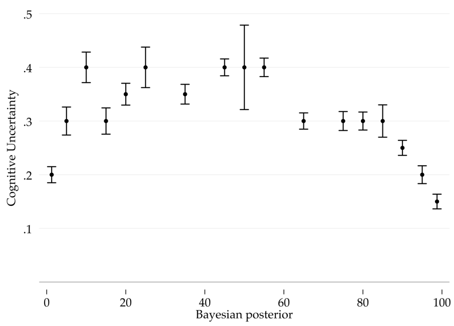
Fits with Enke and Graeber's findings that cognitive uncertainty correlates with regressive-ness:

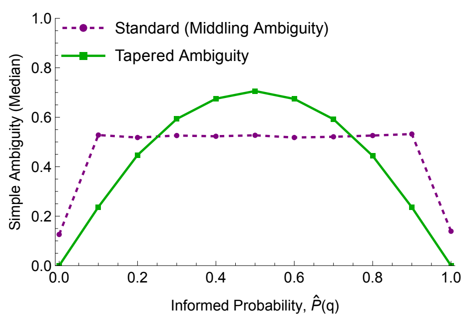


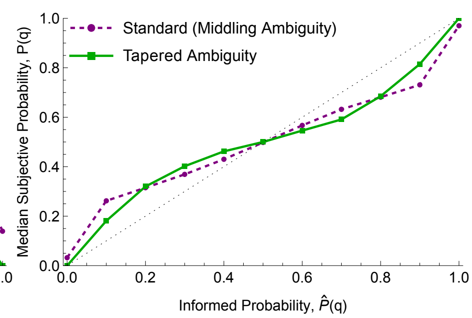
What do the regression coefficients mean? Slope of the line of best fit:



Simulations also replicate inverse-U shape of cognitive uncertainty against informed probability:







Empirical strategy:

- Use simulations to predict correlations between ambiguity and the target effect (hindsight bias, confirmation bias, polarization, etc.)
- Put real people in analogous experimental conditions.
- Use Enke-and-Graeber-style questions to measure ambiguity (cognitive uncertainty).
- See whether results match (qualitative) predictions.

References

Buchak, Lara, 2013. *Risk and rationality*. Oxford University Press.

Quiggin, John, 1982. 'A theory of anticipated utility'. *Journal of Economic Behavior and Organization*, 3(4):323–343.

Tversky, Amos and Kahneman, Daniel, 1992. 'Advances in prospect theory: Cumulative representation of uncertainty'. *Journal of Risk and uncertainty*, 5:297–323.