Noisy Uncertainty: Chapter 2

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

I. Preliminaries

The Challenge: How could you (1) truly have exact[1] probabilities that (2) influence your actions without (3) being able to remove your higher-order uncertainty (**HOU**) by acting?

[1] Or relatively exact

Internalist, behavioralist impulses.

Subjective vs. rational vs. objective probabilities.

Standard operationalization of subjective probabilities: betting.

How much willing to pay for a bet $B =$ \$1 if heads, \$0 if tails?
$P(q) = 0.5 \rightsquigarrow \$0.50$
$P(q) = 0.7 \rightsquigarrow \$0.70$
$P(q) = 0.2 \rightsquigarrow \$0.20$

But then just see what you want to pay for a bet!
$\rightsquigarrow$ So most people working on higher-order uncertainty focus on *rational* probabilities.

E.g. Williamson 2000, 2008; Elga 2013; Lasonen-Aarnio 2015; Salow 2018; Hedden 2019; Dorst 2020; Gallow 2021; Das 2022, etc.

Basic idea:

If you have subjective probabilities, then there are facts about what they are at a given time. They can be objects of uncertainty.

At least for others.

You do have some higher-order opinions about your own opinions.

at least 1 spoon?
(?)$\Rightarrow$ To avoid HOU, must be HO *certain* (**HOC**).

Cognitive noise = randomness between mental state and its effects. Can make mental state difficult to discern.

Example(?): stress vs. dehydration.

Internalist objection: that's uncertainty about the *cause* of the mental state, not the state itself.

Alternatives? Maybe:
1) Appearance of tree's height.
2) Strongest belief about height.
3) How many cookies you want.
4) Which image is clearer (eye doctor).
...

II. Sampling model (Urns in the head)

Sampling is *algorithmic* level, not computational, in Marr's levels.
Levels can interact (e.g. resource constraints; algorithmic complexity).

Likewise, maybe: knowing algorithmic level is noisy can affect whether computational level is HOC or HOU.

Generative models; sampling propensities; elicitations.
Spoon + cereal model.

Vectors as probabilities;
$P = (0.35, 0.15, 0.20, 0.30)$ vs
$\epsilon = (0.27, 0.18, 0.18, 0.36)$

Sampling View of Credences: Your credence in q at time t is the sampling propensity of your generative model to generate q-samples at t.

Separate (1) thing that directly encodes/updates beliefs (credences) from (2) thing that directly interfaces with action (elicitations).

Theoretical evidence: this is how computer scientists have to do it.

Empirical evidence.
- Generative models: intuitive physics and social reasoning

- Probability-matching, 3-cup task, sigmoid response rates.

Example: hard-to-discern generative model. Can condition it.
Precise sampling propensities, but samples always noisy.

Q: Does this match the philosophers' notion of credences?

III. Sampling and *higher-order* uncertainty

How represent that uncertainty explicitly?

Third-person case: they have a distribution over both a 'first-order' question (d or $\neg d$) and a 'second-order' one ($P(d) = 0.55$ or $P(d) = 0.45$).
$\rightarrow$ E.g. Stan: $(0.45, 0.15, 0.10, 0.30)$ over $(d_h, \overline{d}_h, d_l, \overline{d}_l)$.
Encoded in sampling propensities.

More generally, it'll be over a 'P-closed' subject-matter $\mathcal{Q}$, such that the answer to $\mathcal{Q}$ also tells you what P's credences are about $\mathcal{Q}$.

If he observes you elicit your credence by guessing, then he can update on that to get (inconclusive) evidence about what your credences were.

E.g. if know will draw 11 samples and guess most likely, Binomial(11,$P(d)$). (If unsure, marginalize.)

First-person case: What about the *higher-order* case, where you're uncertain about *your own* sampling propensities? I.e. your sampling propensities are uncertain about themselves.

Urn unsure about *its own contents??*

Claim: works the same way. Nothing mysterious.
Updating on your own elicitations still fails to eliminate HOU.

I suspect some self-knowledge people will be unhappy here...

IV. Noisy sampling requires HOU

Standard view: sampling explains ambiguous judgments, not clear ones.

Icard 2016, etc.

But there's a continuum: Maggie-n cases.

Sampling model explains this, once we use your HO opinions.

Outcome- vs. estimate-sampling. *Note:* both can be seen as getting an elicited distribution ϵ, but then *doing* different things with it: calculating $\mathbb{E}_\epsilon(\mathbb{1}_q)$ or $\mathbb{E}_\epsilon(P(q))$.

<u>Argument:</u>

(P1) Sampling: (If sampling model is right, then) if $P(p) = y$, then if you draw a sample about p it's y-likely to be a p-sample.

(P2) Mean elicitations: (If sampling model is right, then) an available strategy for estimating the value of a variable X is to elicit samples and then report the mean value of X amongst those samples.

Conclusion: If sampling model is right, then: if noise is inevitable and you have HO opinions, you must have higher-order uncertainty.

And since *reasonable* agents would avoid noise if they can: if reasonable agents have noise, they must have higher-order uncertainty

<u>Worries:</u>

W1: Sub-personal models with no higher-order opinions?
W2: Why doesn't this imply (falsely) that we should be able to noiselessly sample from the distributions in our computers?

(?): No HO opinions. Or if they do, uncertain b/c logically non-omniscient.

W3: Sampling enough removes HOU.

Reply: That *changes* opinions (to informed probabilities).
Counter-reply: Only applies to computationally limited agents.
Counter-counter: Something about expanding pi. Or about being unsure whether can do iid sampling. Or...

Simulations and Experiment:

Simple idea. Not very diagnostic if succeeds, but *really bad* if it fails.

Simulations: random Ambiguous-Bayesian models [hang tight] with a random variable you have ambiguous estimate about; give Lo, Med, or Hi information about; elicit estimate-samples and track (1) mean elicited estimates and (2) within-agent noise in elicited estimate.

Experiment: Try to do the same thing with real people.

- Tell about WC tasks; repeatedly give scenarios where character X has seen n, and we're tallying how well they did. How many do you estimate they completed?
- 8 scenarios; 4 fillers; 4 targets that varied n and the character X, ($n = 100, 126, 72, 44$).
- Randomize clarity into Lo, Med, and Hi.
- Prediction 1: not much effect on mean estimates.
- Prediction 2: big effect on intra-personal noise in estimates.

To force re-estimating

Lo: 0–100%; Med: 25–75%; Hi: 46–54%

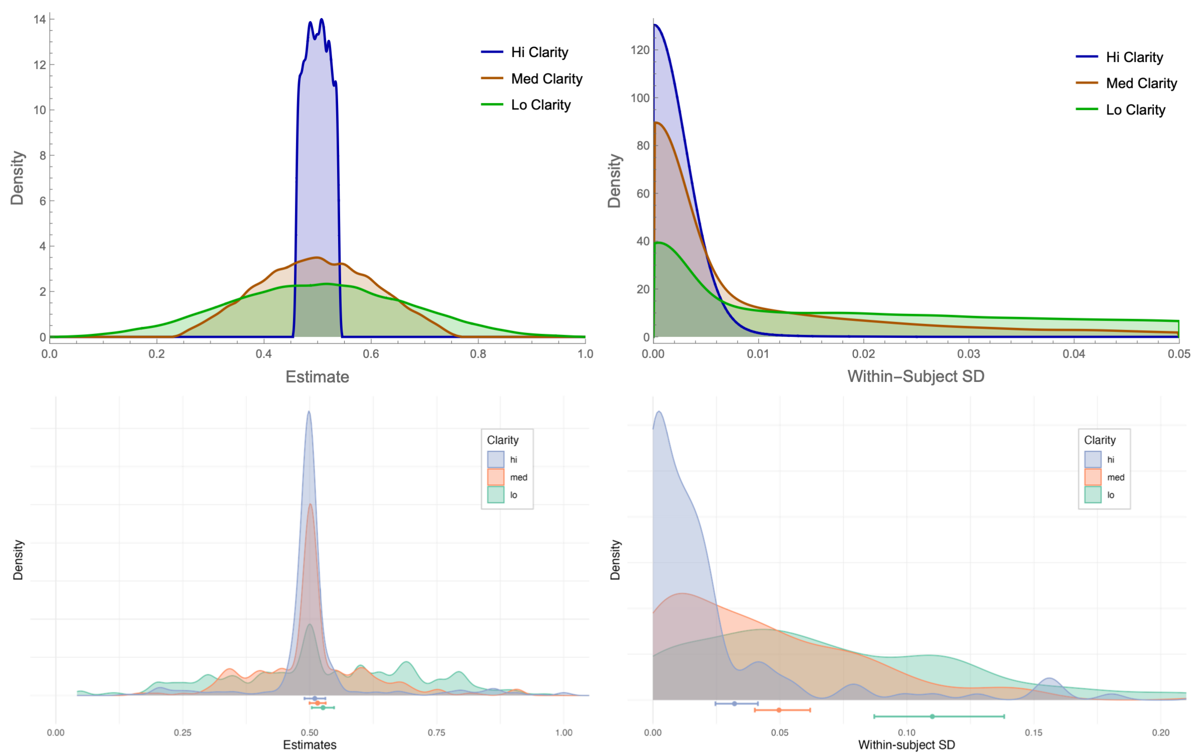

Figure 2.7: **Top:** Simulation results showing (left) individual estimates and (right) within-subject standard deviation in their estimates, divided by clarity. **Bottom:** Experimental results showing the same thing, along with Bayesian-regression estimates and 95%-credible intervals for the means below the x-axis.

V. Generalizing: Cognitive noise implies HOU

Easy to imagine alternatives; eg Gaussian noise in elicited x.

Argument 1: Timothy Williamson says so.

And hence so does Alex Byrne.

Argument 2:

(P1) Cognitive Noise: In estimating your credence in d under ambiguity (in normal conditions), if $P(d) = x$, then there's a $y > 0$ such that there's a decent ($\geq 20\%$) chance that you'll give an estimate for $P(d)$ that's either above $x + y$ or below $x - y$.

With $\mathcal{E}$ a variable for your elicitation: $ch(\mathcal{E} < x - y \text{ or } \mathcal{E} > x + y) \geq 0.2$

(P2) Certainty Precludes Chance: In normal conditions, if you're certain that X is between x_1 and x_2, then you can elicit an estimate for X in a way that has a high ($\geq 90\%$) chance of being b/t x_1 and x_2.

Conclusion: If under normal conditions you have higher-order opinions and inevitably have (significant) noise in your estimate, then you have higher-order uncertainty.

Q: Every time I give this argument it feels too easy. Am I missing something?

Argument 3: Irrelevant influences on confidence rating (Matthias).

Worries:

W4: Imprecision? Bah!

W5: ??

References

Das, Nilanjan, 2022. 'Externalism and exploitability'. *Philosophy and Phenomenological Research*, 104(1):101–128.

Dorst, Kevin, 2020. 'Evidence: A Guide for the Uncertain'. *Philosophy and Phenomenological Research*, 100(3):586–632.

Elga, Adam, 2013. 'The puzzle of the unmarked clock and the new rational reflection principle'. *Philosophical Studies*, 164(1):127–139.

Gallow, J. Dmitri, 2021. 'Updating for Externalists'. *Noûs*, 55(3):487–516.

Hedden, Brian, 2019. 'Hindsight Bias is Not a Bias'. *Analysis*, 79(1):43–52.

Lasonen-Aarnio, Maria, 2015. 'New Rational Reflection and Internalism about Rationality'. In Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology*, volume 5, 145–171. Oxford University Press.

Salow, Bernhard, 2018. 'The Externalist's Guide to Fishing for Compliments'. *Mind*, 127(507):691–728.

Williamson, Timothy, 2000. *Knowledge and its Limits*. Oxford University Press.

———, 2008. 'Why Epistemology Cannot be Operationalized'. In Quentin Smith, ed., *Epistemology: New Essays*, 277–300. Oxford University Press.