

Reasonably Polarized: A Bayesian Theory of Bias

24.805 Topics in Theory of Knowledge

Fall 2025

Lecture: W, 12–3

Classroom: 32-D831

Course website: https://www.kevindorst.com/fa25_RP.html

Professor: Kevin Dorst

kmdorst@mit.edu

Office: 32-D938

Office Hours: Th 4–5, by appt.

Class Description:

This seminar will be about the following thesis:

Reasonable Convergence: If evidence is plentiful, and people are reasonable, they'll (usually) converge to the truth.

Though optimistic about rationality, this thesis is pessimistic about humanity: given the pervasiveness of polarization, it implies that many people are being unreasonable about many things. Assuming that Bayesian convergence theorems have established Reasonable Convergence, behavioral scientists have catalogued a range of biases that they think are symptoms of our irrationality.

This course will center around my book manuscript. It gives a theoretical critique of Reasonable Convergence, and develops it into an empirical theory of the (rational) causes of bias and polarization. The key observation is that the (Bayesian) arguments for Reasonable Convergence implicitly and essentially presuppose **clarity**—that we know exactly what our own opinions are. *Sometimes* our opinions are clear—How likely do you think this fair coin is to land heads? But usually they're **ambiguous**, in the sense that we have uncertainty about what our own subjective probabilities are—How likely do you think it is that I own a dozen spoons? Under ambiguity so defined, Bayesian conditioning often doesn't converge to the truth. I argue that this reconciles the explanatory successes of Bayesian cognitive science with the empirical biases made (in)famous by the research programs of heuristics and biases and behavioral economics.

The book has three main parts. The theoretical part develops the conceptual and mathematical tools to model Bayesians who are unsure what their own subjective probabilities are. The empirical part uses these models to build and test an Ambiguous-Bayesian theory of probability-weighting, hindsight bias, confirmation bias, polarization, and overconfidence. The political part asks how it should affect our approach to psychology and politics if we came to think that people generally form their opinions reasonably—even though, due to ambiguity, many of them are wrong.

Along the way, we'll read related literature in philosophy and cognitive science. Students will have the option to devote some class sessions to tutorials on the methods used throughout the book, including performing experiments and running simulations. No formal tools will be presupposed. Plus you'll get to pick apart a work in progress, and help me make it better!

Readings: All readings on the course website: https://www.kevindorst.com/fa25_RP.html

Requirements:

- Do all required reading. (Optional readings are there to give you places to find paper topics.)
- Attend and participate in all classes.
- Give a short presentation on your idea for a term paper in class on **December 10**.
- Submit a draft of your term paper by **December 16th** (11:59pm).
- Submit a final, revised paper by **January 30, 2026** (11:59pm).

Disability Accommodations: If there's any issue to touch base about, let me know. Disability and Access Services: <https://studentlife.mit.edu/wellbeing-support/disability-and-access-services/>

(Tentative) Schedule

1. **(Sep 3) Reasonable Convergence?**
Chapter 1, The Assumption of Convergence
Mandelbaum 2019, “Troubles with Bayesianism”
Optional: Fine 2005, “The Pigheaded Brain” from A Mind of Its Own
2. **(Sep 10) Cognitive Noise**
Chapter 2, Noisy Uncertainty
Optional: Icard 2016, “Subjective Probability as Sampling Propensity”
3. **(Sep 17) Hindsight Bias**
Chapter 3, What You Need to Know
Hedden 2019, Hindsight Bias is Not a Bias
Optional: Enke and Graeber 2023, “Cognitive Uncertainty”
4. **(Sep 24) Hindsight Bias**
Chapter 7, Hindsight Bias
Optional: Gerken 2024, “Assessing the Evidence for Outcome Bias and Hindsight Bias”
5. **(Oct 1) Possibly: tutorial on running experiments**
6. **(Oct 8) Standard Bayesianism**
Chapter 4, Standard Bayesianism and its Limits
Isaacs and Russell 2022, “Updating Without Evidence”
Optional: Zhang and Meehan, “Bayes is Back”
7. **(Oct 15) Ambiguous Bayesianism**
Chapter 5, Ambiguous Bayesianism
Optional: Elga 2013, “The puzzle of the unmarked clock...”
8. **(Oct 22) Confirmation Bias**
Chapter 8, Confirmation Bias
Optional: Kelly 2008, “Disagreement, Dogmatism, and Belief Polarization”
Optional: McWilliams 2021, “Evidentialism and Belief Polarization”
9. **(Oct 29) Possibly: tutorial on running simulations**
10. **(Nov 5) Rationality**
Chapter 6, The Value of Rationality
Optional: Das 2023, The Value of Biased Information
11. **(Nov 12) Polarization**
Chapter 9, Polarization
Optional: Assaad and Hahn 2024, “Rational Polarization: Sharing Only...”
Optional: Lederman 2015, “People with Common Priors Can Agree to Disagree”
12. **(Nov 19) Mechanistic vs. Rational Psychology**
Chapter 11, The Science of Irrationality
Chater et al. 2024, “A Bayesian Conversation” in *Bayesian Models of Cognition*
Optional: Williams 2021, “Socially Adaptive Belief”
13. **(Nov 26) No class (Thanksgiving)**
14. **(Dec 3) Politics**
Chapter 12, The Politics of Irrationality
Flowerree 2023, “When to Psychologize”
15. **(Dec 10) Student Presentations**