

24.223 Rationality

Fall 2024

Lecture: T/Th 2:30–4

Classroom: 35-308

Course website: kevindorst.com/fa24_rationality

Professor: Kevin Dorst

kmdorst@mit.edu

Office: 32-D938

Office Hours: T 4–5

Class Description: How rational are people? Some bits of cognitive science suggest we’re surprisingly foolish; others suggest we’re amazingly smart. What should we make of this?

Answering this question requires combining both normative and descriptive methods: we need to know how *rational* people would reason, as well as how real people *do* reason. So we’ll learn the basics of Bayesian theories of rationality in order to help us interpret empirical results that have been taken to demonstrate human irrationality. People exhibit the conjunction fallacy, hindsight bias, the gambler’s fallacy, ‘overconfidence’, wishful thinking, confirmation bias, and polarization. Are these evidence for irrationality? We’ll try to figure it out.

Course Goals: Addressing these questions involves combining empirical results, mathematical reasoning, and philosophical reflection. There’s an art to doing this well. We’ll try to learn it.

Readings:

Textbook. You will need **both** the **first** and **second** volumes of Titelbaum, *Fundamentals of Bayesian Epistemology*. MIT gives free online access to the book—you can find it by searching on the [MIT libraries website](#).

Papers. All other readings are posted in a zip file on the course website: kevindorst.com/fa24_rationality

Grading: Your grade will be determined by attendance and participation (10%), an argument summary (10%) and two papers (40% + 40%).

Attendance/participation (10%). Attendance is mandatory. This will be a partially discussion-based class, so to receive full credit you must also contribute at least semi-regularly. If you have trouble speaking up in class, you can email me questions beforehand.

Argument summary (10%). Your first written assignment will be a 2–3 page (600–900-word) summary of an argument in one of the readings. It’s a practice-run for what you need to do in the first half of your papers. **Submitted via email.**

Papers (80%). You will write two papers, each roughly 2000-words (7–8 pages) in response to one of the readings. Each is worth 40% of your grade. **Submitted via email.**

Writing philosophy papers is a learned skill that can be tough to get the hang of. It requires making a single contribution to an ongoing conversation started by the readings and in class. We’ll discuss what’s expected; writing advice is on the website. Prompts will be distributed.

IMPORTANT: ChatGPT and AI Policies.

You are free to use AI models in the development and brainstorming of your ideas—just as you would talk through your paper ideas with a friend, or use a search engine to learn about a topic. **However, you must write and edit your paper yourself. Using AI to write (parts of) your paper for you will count as plagiarism.**

The reasoning: You may never again need to come up with your own arguments or ideas *purely* from scratch, without AI assistance. You certainly **will** need to learn how to develop, formulate, frame, and express your own ideas and arguments in your own words. That is the skill that this course is trying to teach.

The method: Draftback. To help maintain your willpower in the midst of the semester, **you are required to write your entire paper—from start to finish—in Google Docs.** (You can outline and brainstorm in whatever application you like.) When you submit the final version, you will also **share a Google Doc link AND give me editing privileges.** This will allow me to use the “Draftback” Chrome plugin to verify that the paper was written by a human. (It creates a time-lapse of the keystroke-by-keystroke writing of the paper, allowing me to quickly verify that it involved all the typos, false starts, and sentence- and paragraph-rewrites characteristic of human writing.) I will only check the Draftback time-lapse if I have independent reason to suspect that the paper was written by an AI. No part of your writing process will be evaluated.

Failing to submit a Google-Doc link will provide (Bayesian) evidence that the paper is AI-generated. **If it is AI-generated, you will automatically fail the course and be referred to the academic integrity office.** It’s not worth it.

Course Policies:

Announcements. I’ll send class communications via email.

Late work. Late work will be marked down 1/3 of a letter grade for each day late (a B+ becomes a B, etc.).

Writing Help. Clear and concise writing is an important component of this class. The MIT Writing and Communication Center (WCC) offers free one-on-one professional advice from communication specialists with advanced degrees and publishing experience. The WCC can help you learn about all types of academic and professional writing and further develop your oral communication skills. You can learn more about WCC consultations at <http://cmsw.mit.edu/writing-and-communication-center> and register with the online scheduler to make appointments through <https://mit.mywconline.com>. WCC hours are Monday-Friday, 9am–6pm during the semester, and fill up fast.

Academic Integrity. Any writing submitted for a grade must be your own; quotations or ideas paraphrased from other sources should be clearly cited in a way that allows us to find the source. You are responsible for knowing what counts as plagiarism. See <https://integrity.mit.edu/sites/default/files/images/AcademicIntegrityHandbook2020-color.pdf>. Reach out if you have questions. Asking beforehand is never an issue.

Disability Accommodations. MIT is committed to the principle of equal access. Students who need disability accommodations are encouraged to speak with Disability and Access Services (DAS), prior to or early in the semester so that accommodation requests can be evaluated and addressed in a timely fashion. If you have a disability and are not planning to use accommodations, it is still recommended that you meet with DAS staff to familiarize yourself with their services and resources. Please visit the DAS website for contact information: <https://studentlife.mit.edu/das>. If you’ve already been approved for accommodations, please inform me ASAP so we can make an implementation plan.

Schedule

1. (Sept 5) Gershman 2021, pp 1–6
2. (Sept 10) Degrees of belief, Titelbaum Ch. 1
3. (Sept 12) (Conditional) Probability, Titelbaum Ch. 2–3
4. (Sept 17) Conditionalization, Titelbaum Ch. 4 *and* Assad and Hahn 2024 (just get the gist)
5. (Sept 19) The Conjunction Fallacy, Tversky and Kahneman 1983
6. (Sept 24) The Conjunction Fallacy, Dorst and Mandelkern 2022 (skip §3.3, 4.1, and 4.2)
7. (Sept 26) Hindsight bias, Hedden 2019
8. (Oct 1) Hindsight bias, Dorst_HB blog posts (Office hours canceled)
Argument-summary prompts distributed.
9. (Oct 3) Deference Principles, Titelbaum sections 5.1–5.3
10. (Oct 8) The law of small numbers, Tversky and Kahneman 1971
11. (Oct 10) The gambler’s fallacy, Dorst_GF
Argument summary due at 11:59pm. Share google doc permissions.
12. (Oct 15) No class (student holiday)
13. (Oct 17) Hierarchical Models, Kemp et al. 2007
Paper 1 prompts distributed
14. (Oct 22) Overconfidence, Dorst_OC
15. (Oct 24) Overprecision, Moore et al. 2015
16. (Oct 29) Decision Theory, Titelbaum Ch. 7
17. (Oct 31) Motivated Ignorance, Kinney and Bright 2022
18. (Nov 5) Wishful thinking, Melnikoff and Strohminger 2023
Paper 1 due at 11:59pm. Share google doc permissions.
19. (Nov 7) No class (Kevin traveling)
20. (Nov 12) Polarization, Nielsen and Stewart 2018
Paper 2 prompts distributed
21. (Nov 14) No class (Kevin traveling)
22. (Nov 19) Polarization, Kelly 2008
23. (Nov 21) Polarization, Henderson and Gebharter 2021
24. (Nov 26) Confirmation bias. Hahn and Harris 2014, pages 59–76
25. (Nov 28) No class (Thanksgiving)
26. (Dec 3) Confirmation bias, Salow 2018 (section 1) and Dorst_RP (sections 1–2)
27. (Dec 5) Polarization, Dorst_RP, sections 3–8
28. (Dec 10) Selective memory, Wilson 2014
29. (Dec 11) No class. **Paper 2 due at 11:59pm. Share google doc permissions.**