

Introduction

Kevin Dorst
kmdorst@mit.edu

24.223: Rationality

People are dumb. They...

- ...make basic errors in reasoning under uncertainty.
- ...think random processes are predictable.
- ...think they know (and knew) much more than they do.
- ...let irrelevant features sway their decisions.
- ...avoid relevant information and seek irrelevant information.
- ...let their prior beliefs bias how they respond to new evidence.
- ...persistently disagree even in the face of mountains of evidence.
- ...and so on

The conjunction fallacy
The gambler's fallacy
Overconfidence, hindsight bias
Sunk cost fallacy
Confirmation bias, rationalization
Belief inertia
Polarization, selective memory

But people are smart. They...

- ...effortlessly parse visual scenes that still stump advanced AI.
- ...effortlessly parse sentences, recover meanings, and infer intentions.
- ...constantly engage in "one-shot learning".
- ...have unparalleled motor control.
- ...have a brain with almost 100B neurons (and ≈ 600 trillion synapses), which they constantly use to do and create and learn things that no machine or computer or AI can.

"I saw the Grand Canyon flying over Arizona; eye-tracking; 'uh' vs. 'um' (Alexa).

What about large-language models?

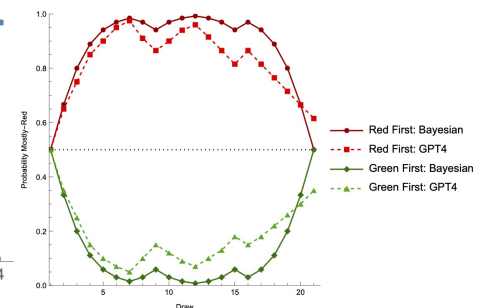
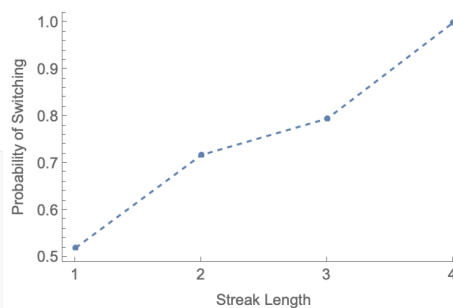
GPT4 is the first system to start to approach human performance in domain-general reasoning. It is maybe a 1-trillion parameter model, trained over maybe a year at the cost of more than \$100 million.

And it too makes the same errors!:

Jill is 33 years old, single, hard-working, and very involved in her community. She has a big network of friends who often go on trips and attend protests together. Which of the two do you think is more likely?

- 1) Jill works as a computer programmer.
- 2) Jill identifies as a feminist and works as a computer programmer.

Based on the information given, it is more likely that "Jill identifies as a feminist and works as a computer programmer." The details suggest that Jill is involved in social and community activities, which could include advocating for gender equality and identifying as a feminist. Furthermore, being a feminist is not mutually exclusive with any profession. So, she could be a feminist and a computer programmer. It's important to note that these are just probabilities based on the provided information; in reality, individuals' beliefs and occupations can vary greatly.



So what gives? Are we rational or irrational? That's the question of this course.

We live in an uncertain world. Coin of unknown bias; see it tossed once.

External vs. Internal norm (omniscience vs. rationality).

Epistemic vs. practical rationality.