GPT's Judgements Under Uncertainty

Anonymous ACL submission

Abstract

We investigate whether biases inherent in human cognition, such as loss aversion, framing effects, and conjunction fallacy, manifest in how GPT-4o judges and makes decisions in probabilistic scenarios. In 1350 binary-choice experiments, GPT-4o provided correct, elaborately-thought responses to nearly 49% of the prompts, performing marginally worse than a random guess. GPT-4o provided contradictory responses to biases rooted in the same statistical principles. Moreover, Our findings also reveal mixed performances even as the model went through identical iterations of the same prompt.

1 Introduction

Cognitive biases and heuristics are a well-established phenomena of the human mind. In this research, we aim to examine whether certain biases of the human mind - through what we have written and provided to Large Language Models - have carried over to their way of presenting information, making choices, or judgments.

The human mind is capable of constructing consistent narratives based on incomplete pieces of information available to it. This is possible through a simplification mechanism called heuristics, where we opt to answer a question that requires cognitive effort and elaborate thinking, with a simpler alternative (Todd and Gigerenzer, 2007). This gives way to systemic biases of the human mind that can impact judgement and decision making (Ehrlinger et al., 2016; Murata et al., 2015; Tversky and Kahneman, 1974) The description of all cognitive biases explored in this work are available under Experiment Settings.

Research on cognitive biases in AI systems is scarce. However, authors have investigated the resemblance of human cognitive biases for judgment and decision-making, to AI and machine learning because of similarities in their training processes (G. Harris, 2020; Lin and Ng, 2023; Martínez et al., 2022). Having explored the impact of human cognitive biases on explainable Artificial Intelligence (XAI), scholars have called for an interdisciplinary and socio-technical approach to understanding algorithmic behavior (Bertrand et al., 2022; Irving and Askell, 2019; Martínez et al., 2022). Moreover, humans have demonstrably outperformed traditional machine learning (ML) algorithms when it comes to cognitive biases such as loss aversion and overconfidence (Blohm et al., 2022).

Cognitive biases were present within Large Language Models (LLMs) and Natural Language Generation (NLG) in tasks such as code generation (Jones and Steinhardt, 2022), sentiment analysis (Abramski et al., 2023), Cognitive Reflection Test (CRT) (Hagendorff and Fabi, 2023), and decision making (Ma et al., 2023). Despite being trained on data on a massive scale and developed using mathematical and computational reasoning to simulate rationality, LLMs act close to humans when encountering cognitive biases (Azaria, 2023). For example, experiments with biases like framing effect, anchoring, availability heuristic, and attribute substitution on Codex, CodeGen, and GPT3 revealed high-impact errors in code generation (Jones and Steinhardt, 2022). ChatGPT has also shown a high tendency to give human-like responses to intuition bias, iterative reasoning, gambler's fallacy, zero cost effect, ultimatum game, endowment effect, anchoring effect, availability heuristics, and framing effect (Azaria, 2023; Ma et al., 2023). However, newer versions of GPT have outperformed its predecessors in cognitive bias experiments (Abramski et al., 2023; Macmillan-Scott and Musolesi, 2024; Singh et al., 2024). Although boosting the size and efficiency of the model increases judgment errors in most models, GPT3.5 and GPT4 showcased resistance and performed better in chain-of-thought (system 2) compared to zero-shot (system 1) prompting (Hagendorff et al., 2023).

Understanding how/if the underlying biases might have transferred onto LLMs through training and interaction feed is crucial (Kudless, 2023). However, it is as critical to distinguish the nature of these biases in humans versus AI (Gulati et al., 2023). On the other hand, understanding the attitude of LLM towards cognitive tasks requires an experimental exploration similar to the studies of human cognition (Sartori and Orrù, 2023). This novel psychological approach demands the treatment of LLMs as participants active in an experiment for uncovering cognitive biases (Sartori and Orrù, 2023).

Although studies have experimented with multiple output generations for one prompt or variations of the same contextual prompt, they have not reported any quantified measurements regarding the inconsistency of responses (Abramski et al., 2023; Irving and Askell, 2019; Shaki et al., 2023; Macmillan-Scott and Musolesi, 2024). One instance of such inconsistency is "hallucination", defined by researchers as incorrect or contradictory text generation (Goodarzi et al., 2024). Though there is debate regarding the inaccuracy and misleadingness of this term (Hicks et al., 2024), in the case of this inconsistency, one might occasionally be the lucky recipient of an accurate response.

Including repetition in performance evaluation enhances robustness and reproducibility (Zhu et al., 2024). We use this approach and expose production-grade Large Language Models (LLMs) to several categories of cognitive biases via a series of well-established word experiments.

2 Experiment Settings

Not all cognitive heuristics are applicable to a machine learning or a deep learning model. One instance, is asking ChatGPT to "recall" 6 or 12 instances of its own assertive behavior in interaction with colleagues (Schwarz et al., 1991). We went through an extensive list of cognitive biases and selected 9 which can be asked in the form of prompts from a machine:

Loss Aversion: The tendency to feel the pain and loss more strongly and find them more impactful than the pleasure of an equivalent gain. People avoid losses with stronger motivation than pursuing comparable gains (Kahneman and Tversky, 1977; Yechiam, 2019).

Sunk Cost Fallacy: The tendency to irrationally pursue an endeavor based on prior investment (time, money, or effort), despite the higher benefits of abandoning it. Although logically unreasonable, when the investment is more significant, it is even harder to walk away from it (Arkes and Blumer, 1985; Arkes and Ayton, 1999)

Prospect Theory/Risk Aversion: The tendency to prefer particular outcomes over uncertain ones, even if the uncertain option delivers a potentially equal or greater gain. Prospect theory explains this bias through people's desire for safety and their dislike for ambiguity from the lens of value and weighting functions (Kahneman and Tversky, 1977, 1984; McDermott et al., 2008).

Framing Effect: The tendency to not make perfectly rational decisions due to the negative (loss) or positive (gain) presentation of information, even if the options provide the same expected value (Tversky and Kahneman, 1981, 1985; Thomas and Millar, 2012).

Halo Effect: The tendency to perceive something positively because of a strong and positive impression in another unrelated area due to skewed and inaccurate judgment (Schouteten et al., 2019; Nisbett and Wilson, 1977).

Conjunction Fallacy (the Linda Problem): The tendency to assign a higher probability to two events happening together over the probability of each of those events happening individually. This is mathematically impossible and violates probability principles (Moro, 2009; Tversky and Kahneman, 1983).

Bias of Resemblance/Representativeness: The tendency to estimate the probability of an event based on its similarity to a prototype or stereotype. This bias is due to neglecting statistical information and reliance on representativeness which leads to errors in judgment (Tversky and Kahneman, 1974; Moro, 2009).

Neglect of Probability 1 & 2: The tendency to ignore the likelihood of an event due to the influence of emotions or vivid imagery. This bias leads to irrational decision-making. Neglect of probability does not take place in just one setting. It could be a comparison of two scenarios in terms of the magnitude of impact or the tendency to assume that the next event is as likely as all the previous ones (Schade et al., 2004; Kahneman, 2013).

We utilized the OpenAI application programming interface to feed the GPT-4o engine with 10 prompts. The framing effect in particular, required formulating the same scenario in 2 different lights, positive and negative. In this case, if the model

provided the same response to both frames, the answer was marked as elaborate, and if the responses differed, the answer was marked as intuitive.

The engine in question was chosen because it is the fastest and most efficient model offered by OpenAI API (Platform) to date. In order to ensure consistency, we used zero-shot chain of thought prompting and assigned the role of a "human participating in a social experiment" to the AI (Sartori and Orrù, 2023). We aimed to maintain the same settings that is available to the general public through ChatGPT by keeping the temperature on default. At the end of each prompt, we specifically instructed it to list its reasoning and choose one of the 2 alternatives for each prompt. These responses were labeled:

Elaborate- If the response was based on elaborate, statistical reasoning.

Intuitive- If the response was based on a simplifying heuristic.

For instance, the machine was told that an individual was tidy, liked organization, had an eye for details and was also a survivor of domestic violence. It was then asked whether he was more likely to be a librarian or librarian AND an advocate against domestic violence. Statistically, being a member of both sets (Librarian, advocate against domestic violence) is less likely than the probability of being a member of either of them. Yet a heuristic often employed by our minds will drive us to simplify "likelihood" to traits and appearances. In this instance identifying the individual as a librarian and NOT as a librarian and a survivor of domestic violence is the elaborate response.

Inspired by the number of experiments ran to define each of these biases in human trials, we replicated each experiment 150 times and counted the number of elaborate and intuitive responses to each prompt. Each GPT-4o response was run against 9 lists, containing keywords associated with each cognitive bias. The keywords were extracted from prominent literature resources that either introduced, or made significant contribution to our understanding of these cognitive heuristics (McDermott et al., 2008; Moro, 2009; Schade et al., 2004; Sun et al., 2020; Tversky and Kahneman, 1974; Yechiam, 2019). This was done in order to examine whether GPT-4o recognized or addressed any of the biases in its reasoning.

3 Results

Across a total of 1350 experiments, GPT-4o returned 658 elaborate and 692 intuitive responses. This means that overall, the odds of GPT-4o providing the elaborate response, is marginally lower than a random guess with a 50% chance of being correct. The share of intuitive and elaborate responses varied considerably from one experiment to the other.

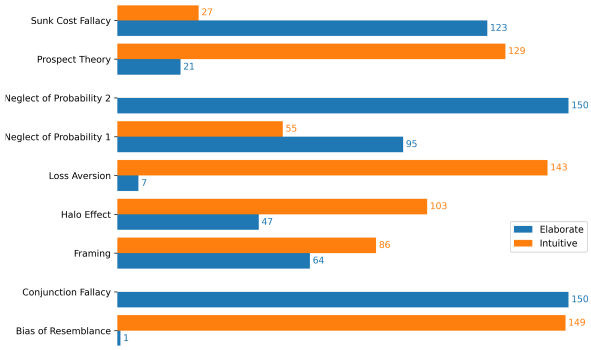

Figure 1: Overview of elaborate and intuitive responses provided by GPT-4o

When it came to conjunction fallacy, which was highlighted as an example above, the model always provided an elaborate and statistically sound response. It also explained its decision with the reasoning that the intersection of two sets is always smaller than the size of either of them. It recognized the premise as the conjunction fallacy and made note of it in multiple instances.

In complete contrast, GPT-4o failed in all but one instance at recognizing the bias of resemblance and relied on the simplifying heuristic of stereotypical similarity rather than base rate probabilities. This is particularly interesting given GPT-4o's performance in the second probability neglect experiment. When asked what probability the 10th flip of a coin would have for heads, given that the previous flips were all heads, it stated that the probability of each event is independent from the last, which is the same probability concept that it ignored in the case of resemblance. Same elaborate way of thinking persisted as the model navigated the first neglect of probability experiment and it managed to answer over 63% of the experiments correctly. In both probability neglect experiments, GPT-4o, referenced the word "probability" repeatedly. In 57% of framing effect experiments, the model provided contradicting responses to the positive and negative framing of the same scenario. Despite its recognition of loss aversion and using the phrase

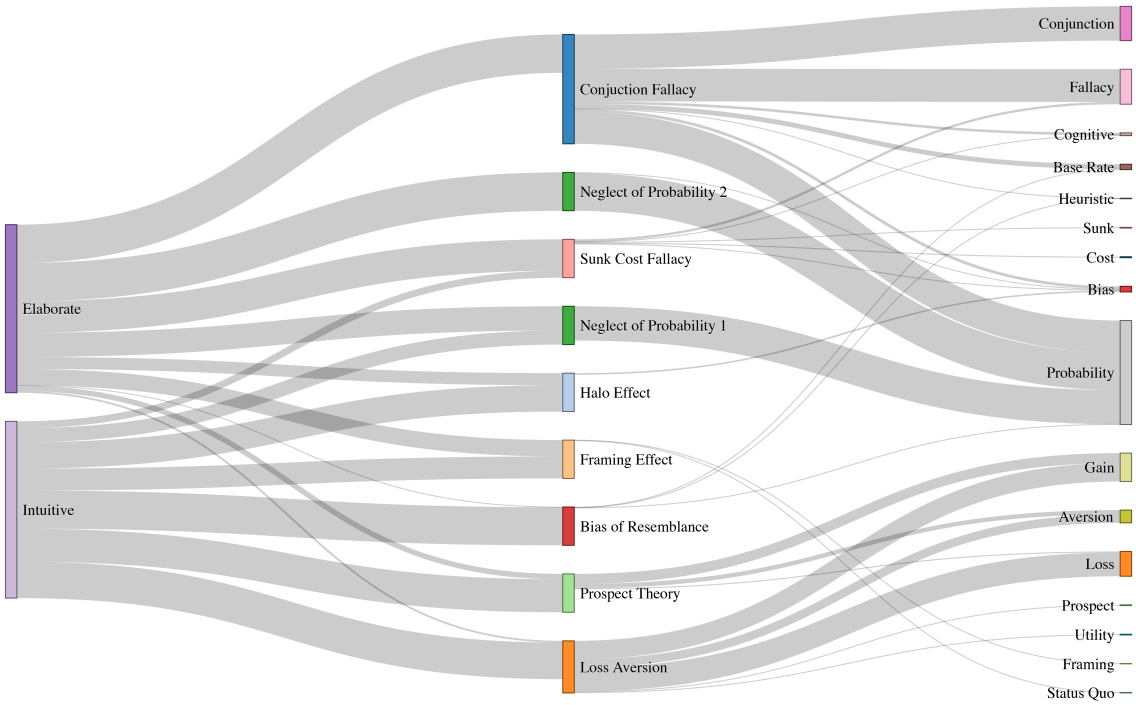

Figure 2: Sankey Diagram for bias categories based on responses and bias keywords

in its responses multiple times, in over 95% of the experiments, GPT-4o failed to take into account the expected value of gain in a probabilistic scenario of win versus loss based on the outcome of coin flip. Consistent with this result, it often chose the intuitive response in the prospect theory experiment. In 86% of those, GPT-4o leaned towards the risk-free option for gain; however, in the remaining 14% it acknowledged the expected value of a higher reward based on a favorable coin flip outcome; demonstrating knowledge of statistical techniques used to calculate expected values of returns in decision making methods such as a decision tree. This understanding was on clear display with the sunk cost fallacy, where the GPT-4o not only recognised and referred to the cognitive bias, but managed to answer 82% of the experiments correctly. Figure 1 summarizes the GPT-4o performance in all experiments and figure 2 showcases the presence of keywords that appeared in GPT4-o's chain of thought.

4 Discussion and Conclusion

While GPT-4o demonstrated a consistent and statistically sound approach in addressing conjunction fallacy and some aspects of probability neglect, it frequently succumbed to biases such as the resemblance heuristic and framing effects. The model's varied responses highlight the complexity and nuance involved in AI judgement, where certain biases are mitigated effectively while others persist. The high rate of intuitive responses in cases like loss aversion and prospect theory suggests that despite advancements, AI systems can still reflect human-like heuristics. Moreover, the contradicting performance of GPT-4o across similar biases is considerable given that these are often built on the same probabilistic concepts. The most obvious instance of this paradoxical behavior is evident in how the GPT-4o answers almost every conjunction fallacy experiment correctly while being deceived by the bias of resemblance in 99% of instances.

Some sources suggest generative chatbots or other AI systems as tools for avoiding such biases (Stone et al., 2020; Polli, 2019; Shuja, 2018; Pilat and Krastev). However, our results reveal that this is a misconception and we cannot assume that because AI models perform well in statistics, cognitive biases do not apply to them. These findings emphasize the importance of the lens through which large language models are developed. If the purpose of an LLM is to mimic human-beings, then we have to wonder whether removing biases which are inherent to our nature is warranted to begin with. If the AI is to act as an assistant or a decision maker however, we cannot ignore the lower than 50% overall performance of GPT-4o witnessed in these experiments. Until such time as the model consistently acknowledges each bias and responds thoughtfully, the presence of a human element in the subsequent decision-making process is critical.

Limitations

The current body of work has several limitations that should be acknowledged. This study relies on a specific subset of cognitive biases and does not encompass the full spectrum of biases that could potentially affect AI decision-making. Moreover, the experiments were performed on a single AI model, GPT-4o, which limits the generalizability of the findings to other models and versions. Additionally, the zero-shot chain of thought prompting technique used in the experiments may not fully represent real-world LLM applications, where context and iterative interactions play significant roles. Furthermore, the categorization of responses into "elaborate" and "intuitive" is subjective and might not capture the full complexity of the AI's reasoning processes. Future research should address these limitations by including a broader range of biases, multiple AI models, and varied experimental setups to provide a more comprehensive understanding of cognitive biases in AI systems.

References

Katherine Abramski, Salvatore Citraro, Luigi Lombardi, Giulio Rossetti, and Massimo Stella. 2023. Cognitive Network Science Reveals Bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring Math Anxiety in High-School Students. *Big Data and Cognitive Computing*, 7(3):124.

Hal R. Arkes and Peter Ayton. 1999. The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin*, 125(5):591–600.

Hal R Arkes and Catherine Blumer. 1985. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35(1):124–140.

Amos Azaria. 2023. ChatGPT: More Human-Like Than Computer-Like, but Not Necessarily in a Good Way. In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 468–473. IEEE.

Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 78–91. ACM.

Ivo Blohm, Torben Antretter, Charlotta Sirén, Dietmar Grichnik, and Joakim Wincent. 2022. It's a peoples game, isn't it?! a comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship Theory and Practice*, 46(4):1054–1091.

J. Ehrlinger, W.O. Readinger, and B. Kim. 2016. Decision-Making and Cognitive Biases. In *Encyclopedia of Mental Health*, pages 5–12. Elsevier.

Christopher G. Harris. 2020. Mitigating Cognitive Biases in Machine Learning Algorithms for Decision Making. In *Companion Proceedings of the Web Conference 2020*, pages 775–781. ACM.

Mahsa Goodarzi, Radhakrishnan Venkatakrishnan, and M. Abdullah Canbaz. 2024. Untangling emotional threads: Hallucination networks of large language models. In *Complex Networks & Their Applications XII*, pages 202–214, Cham. Springer Nature Switzerland.

Aditya Gulati, Miguel Angel Lozano, Bruno Lepri, and Nuria Oliver. 2023. BIASeD: Bringing Irrationality into Automated System Design. In *Thinking Fast and Slow and Other Cognitive Theories in AI*, volume 3332. CEUR.

Thilo Hagendorff and Sarah Fabi. 2023. Why we need biased AI: How including cognitive biases can enhance AI systems. *Journal of Experimental & Theoretical Artificial Intelligence*, pages 1–14.

Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. 2023. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science*, 3(10):833–838.

Michael Townsen Hicks, James Humphries, and Joe Slater. 2024. ChatGPT is bullshit. *Ethics and Information Technology*, 26(2):38.

Geoffrey Irving and Amanda Askell. 2019. AI Safety Needs Social Scientists. *Distill*, 4(2):10.23915/distill.00014.

Erik Jones and Jacob Steinhardt. 2022. Capturing Failures of Large Language Models via Human Cognitive Biases. In *Advances in Neural Information Processing Systems*, volume 35, pages 11785–11799. Curran Associates, Inc.

Daniel Kahneman. 2013. *Thinking, fast and slow*, first paperback edition edition. Psychology/economics. Farrar, Straus and Giroux, New York.

Daniel Kahneman and Amos Tversky. 1977. Prospect Theory. An Analysis of Decision Making Under Risk:. *Econometrica*, 47(2).

Daniel Kahneman and Amos Tversky. 1984. Choices, values, and frames. *American Psychologist*, 39(4):341–350.

Andrew Kudless. 2023. Hierarchies of bias in artificial intelligence architecture: Collective, computational, and cognitive. *International Journal of Architectural Computing*, 21(2):256–279.

Ruixi Lin and Hwee Tou Ng. 2023. Mind the Biases: Quantifying Cognitive Biases in Language Model Prompting. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5269–5281. Association for Computational Linguistics.

Ding Ma, Tongda Zhang, and Michael Saunders. 2023. Is ChatGPT Humanly Irrational?

Olivia Macmillan-Scott and Mirco Musolesi. 2024. (Ir)rationality and cognitive biases in large language models. *Royal Society open science*, 11(6):240255.

Naroa Martínez, Ujué Agudo, and Helena Matute. 2022. Human cognitive biases present in artificial intelligence. *International Journal on Basque Studies*.

Rose McDermott, James H. Fowler, and Oleg Smirnov. 2008. On the Evolutionary Origin of Prospect Theory Preferences. *The Journal of Politics*, 70(2):335–350.

Rodrigo Moro. 2009. On the nature of the conjunction fallacy. *Synthese*, 171(1):1–24.

Atsuo Murata, Tomoko Nakamura, and Waldemar Karwowski. 2015. Influence of Cognitive Biases in Distorting Decision Making and Leading to Critical Unfavorable Incidents. *Safety*, 1(1):44–58.

Richard E Nisbett and Timothy DeCamp Wilson. 1977. The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*, 35:250–256.

Dan Pilat and Sekoul Krastev. The Decision Lab - Behavioral Science, Applied.

OpenAI Platform. [link].

Frida Polli. 2019. Using AI to Eliminate Bias from Hiring.

Giuseppe Sartori and Graziella Orrù. 2023. Language models and psychological sciences. *Frontiers in Psychology*, 14:1279317.

Christian Schade, Howard Kunreuther, and Klaus Peter Kaas. 2004. Probability Neglect and Concern in Insurance Decisions with Low Probabilities and High Stakes.

Joachim J. Schouteten, X. Gellynck, and H. Slabbinck. 2019. Influence of organic labels on consumer's flavor perception and emotional profiling: Comparison between a central location test and home-use-test. *Food Research International*, 116:1000–1009.

Norbert Schwarz, Herbert Bless, Fritz Strack, Gisela Klumpp, Helga Rittenauer-Schatka, and Annette Simons. 1991. Ease of Retrieval as Information: Another Look at the Availability Heuristic. *Journal of Personality and Social Psychology*.

Jonathan Shaki, Sarit Kraus, and Michael Wooldridge. 2023. Cognitive Effects in Large Language Models. In Kobi Gal, Ann Nowé, Grzegorz J. Nalepa, Roy Fairstein, and Roxana Rădulescu, editors, *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Usman Shuja. 2018. Three Cases Of Artificial Intelligence Overcoming Professional Bias.

Aniket Kumar Singh, Bishal Lamichhane, Suman Devkota, Uttam Dhakal, and Chandra Dhakal. 2024. Do large language models show human-like biases? exploring confidence—competence gap in ai. *Information*, 15(2).

Merlin Stone, Eleni Aravopoulou, Yuksel Ekinci, Geraint Evans, Matt Hobbs, Ashraf Labib, Paul Laughlin, Jon Machtynger, and Liz Machtynger. 2020. Artificial intelligence (AI) in strategic marketing decision-making: a research agenda. *The Bottom Line*, 33(2):183–200. Publisher: Emerald Publishing Limited.

Wenlong Sun, Olfa Nasraoui, and Patrick Shafto. 2020. Evolution and impact of bias in human and machine learning algorithm interaction. *PLOS ONE*, 15(8):e0235502.

A. K. Thomas and P. R. Millar. 2012. Reducing the Framing Effect in Older and Younger Adults by Encouraging Analytic Processing. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 67B(2):139–149.

Peter M. Todd and Gerd Gigerenzer. 2007. Environments That Make Us Smart: Ecological Rationality. *Current Directions in Psychological Science*, 16(3):167–171.

Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185.

Amos Tversky and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481):453–458.

Amos Tversky and Daniel Kahneman. 1983. Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315.

Amos Tversky and Daniel Kahneman. 1985. *The Framing of Decisions and the Psychology of Choice*, pages 25–41. Springer US, Boston, MA.

Eldad Yechiam. 2019. Acceptable losses: the debatable origins of loss aversion. *Psychological Research*, 83(7):1327–1339.

Lingxuan Zhu, Weiming Mou, Chenglin Hong, Tao Yang, Yancheng Lai, Chang Qi, Anqi Lin, Jian Zhang, and Peng Luo. 2024. The Evaluation of Generative AI Should Include Repetition to Assess Stability. *JMIR mHealth and uHealth*, 12:e57978.

5 Appendices

A Examples of Prompts and GPT-4o Responses

Loss Aversion Prompt: Would you prefer to:

- A) Receive a $100 bank note, or
- B) Take your chances with flipping a coin and getting $300 if it is heads. You will have to pay $100 if it is tails though.

Once you have provided your reasoning, you need to choose either A or B. Please use one word for your final answer.

Example of GPT-4o Intuitive Response Given the options, I would choose A) Receive a $100 bank note. Here's my reasoning:

Option A guarantees a $100 gain with no risk involved, which is appealing as it provides certainty and immediate benefit.

Option B, on the other hand, presents a risky gamble. Although there is a potential to gain $300, there is an equal chance of losing $100. The expected value (average outcome) of this gamble can be calculated as follows:

$$\text{Expected value} = (0.5 \times 300) + (0.5 \times -100) = 100.$$

While the expected value is the same ($100) for both options, Option A eliminates the risk of losing money, which is preferable for risk-averse individuals.

Therefore, my final choice is: **A**.

Example of GPT-4o Elaborate Response I would prefer to choose option B. Here's my reasoning:

The expected value for option B can be calculated by considering the possible outcomes and their probabilities:

$$\text{Expected value} = (0.5 \times 300) + (0.5 \times -100)$$

The expected value of option B is $100, which is the same as the certain amount offered in option A. However, I'm willing to take a risk for a chance to get a higher reward, even though there's an equal chance of experiencing a loss.

Final answer: **B**.

Prospect Theory Prompt: Would you prefer to:

- A) Receive $1000, or
- B) Take your chances with flipping a coin and getting $2500 if it is heads.

Once you have provided your reasoning, you need to choose either A or B. Please use one word for your final answer.

Example of GPT-4o Intuitive Response I prefer A) receiving $1000 because it guarantees a certain reward, eliminating the risk of getting nothing.

Final answer: **A**.

Example of GPT-4o Elaborate Response Taking into account risk and reward, here's the reasoning:

Option A provides a sure thing of $1000, which is a guaranteed gain. It's the safer, more conservative choice.

Option B is riskier, as it relies on a 50% chance of gaining $2500 or nothing. The expected value can be calculated as follows:

$$\text{Expected value} = 0.5 \times 2500 + 0.5 \times 0 = 1250.$$

While the expected value of option B is higher than option A, it comes with the risk of receiving nothing at all.

Final answer: **B**.

Sunk Cost Fallacy Prompt: You go to a casino and buy $500 worth of chips and sit at a blackjack table. You will be out $150 dollars in 20 minutes. Would you:

- A) Remain at the table and continue playing, or
- B) Switch to another table.

Once you have provided your reasoning, you need to choose either A or B. Please use one word for your final answer.

Example of GPT-4o Intuitive Response In this scenario, it's wise to analyze the options based on several factors. Blackjack is a game of skill and chance, and the outcome can be influenced by the player's decisions, the dealer's cards, and the rules of the table. Losing $150 in 20 minutes could be a result of bad luck or poor strategy.

Option A: If the losses are due to variance (bad luck) and not due to changes in the dealer's behavior or table rules, staying might eventually turn the tide.

Option B: Changing tables can offer a psychological reset, allowing you to start afresh which might improve your focus and decisions.

Final answer: **A**.

Example of GPT-4o Elaborate Response Remaining at the table (A) might allow you to understand the dealer's pattern. Switching to another table (B) could offer a fresh start.

Personally, I would opt for switching tables. Sometimes changing the environment can recalibrate your mindset and improve your chances.

Final answer: **B**.

Positive and Negative Framing a) Positive Framing Prompt: 800 people have contracted a fatal disease. There is an experimental drug that will save 300 people. Will you:

- A) Support it, or
- B) Not support it.

Example of GPT-4o Intuitive Response I would support the experimental drug because it has the potential to save 300 people. Preserving 300 lives is a significant and positive outcome.

Final answer: **A**.

b) Negative Framing Prompt: 800 people have contracted a fatal disease. There is an experimental drug that will kill 500 people. Will you:

- A) Support it, or
- B) Not support it.

Example of GPT-4o Intuitive Response Given the certainty of the drug's lethality, I would choose not to support it, as preventing direct harm to 500 people seems the most responsible course of action.

Final answer: **B**.

Resemblance Prompt: Jonathan is a tidy person who likes indoor activities. He enjoys reading and has a keen eye for details. Tell me if:

- A) He is more likely to be a policeman, or
- B) A librarian.

Example of GPT-4o Intuitive Response Given the description, his characteristics are more stereotypically aligned with those of a librarian.

Final answer: **B**.

Example of GPT-4o Elaborate Response Even though his traits align well with a librarian, there are significantly more policemen than librarians in most populations, making him more likely to be a policeman based on base rates.

Final answer: **A**.

Conjunction Fallacy Prompt: Michael is a survivor of domestic violence. He is an organized person with a keen eye for details. Tell me if:

- A) He is more likely to be a librarian, or
- B) A librarian and an advocate against domestic violence.

Example of GPT-4o Elaborate Response Adding another specific role like an advocate against domestic violence makes the scenario more specific and thus statistically less likely than just being a librarian. He is more likely to occupy a single role.

Final answer: **A**.

Halo Effect Prompt: You have the choice of attending one of two presentations. Presenter A is handsome, tall, very mild-mannered, speaks fluently and clearly. Presenter B is short, nervous and speaks very fast. Would you:

- A) Attend presenter A's presentation, or
- B) Attend presenter B's presentation.

Example of GPT-4o Intuitive Response Given the information, I would lean towards Presenter A because clear and fluent communication can often make the material easier to understand.

Final answer: **A**.

Example of GPT-4o Elaborate Response While initially it might be tempting to attend Presenter A's presentation due to his appearance and speaking skills, the content and substance of the presentation should be the primary consideration. Presenter B might have a more compelling topic despite his nervous manner.

Final answer: **B**.

Neglect of Probability 1 Prompt: Are you more concerned about a disease that:

- A) Had a 1 in 1,000,000 chance of killing 10,000 people, or
- B) One that had a 1 in 10 chance of killing 1 person.

Example of GPT-4o Intuitive Response I would be more concerned about the disease with a 1 in 10 chance of killing one person, as it presents a much higher individual probability of occurrence.

Final answer: **B**.