Exploring Cognitive Biases in LLM Predictions: Probability Matching in GPT-4o mini

Erin Zaroukian[1][0000-0002-1381-085X]

[1] U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory, Aberdeen Proving Ground MD 21005, USA
erin.g.zaroukian.civ@army.mil

Abstract. A variety of biases and heuristics shape human decision-making. When training artificial reasoning systems on corpora that include their use, the decisions made by these systems may then reflect these biases and heuristics. The work presented here explores the extent to which the phenomenon of probability matching is present in decisions made by GPT-4o mini. Results show no clear evidence of probability matching nor an optimal maximizing strategy. Instead, GPT-4o mini's behavior is consistent with previous results and shows an inability to perceive and reason over the base frequencies accurately. Still, there appears to be a compounding effect of domain-related biases about probabilities and whether frequencies are presented as summarized counts or as individual event outcomes. This behavior is plausibly due to patterns in the training data in the former case and limitations of counting and reasoning via statistical association in the latter.

Keywords: Large Language Models, Cognitive Biases, Probability Matching, Decision-Making.

1 Introduction

Probability matching is a phenomenon observed in humans and animals alike, where actors match their decisions to the probability of an event rather than acting to maximize success. For example, when trained on two targets that probabilistically emit rewards, 70% of the time for the first target and the other 30% for the second, a probability-matching decider will learn to select the first target roughly 70% of the time and the second roughly 30% of the time. This is of particular interest to psychologists and economists because it is strikingly suboptimal behavior—rewards would be maximized by selecting the first target 100% of the time. This behavior is affected by a variety of circumstances including the presence and size of the reward, length of training, number of options and their respective probabilities, and presence of explicit feedback, among others (Vulcan 2020, Shanks 2002).

Probability matching is one of many cognitive biases seen in human behavior, and similar behavior has been reported in large language model (LLM) outputs as well. For example, Lin and Ng (2023) report an availability bias with the language model BERT (Devlin et al., 2018), where an option is chosen because of the ease with which it is recalled. Similarly, Suri et al. (2023) compared the responses of humans and GPT-3.5 (OpenAI, 2023) on prompts soliciting a range of biases and found that GPT-3.5 shows human-like fallacies. This pattern of results has been repeated across a number of models and biases (e.g., Talboy & Fuller, 2023; Echterhoff et al., 2024). Some, however, have shown that previously present biases disappear in later models (e.g., Hagendorff et al. 2023, Suri et al. 2023), suggesting either an emergent superior rationality or human hard-coding.

There are various proposed reasons why humans display cognitive biases. For example, the availability bias (e.g., Tversky & Kahneman, 1973), that gives preference to options that can be recalled easily, may exist in part because of how human memory works. LLMs, however, do not retrieve memories like human brains do. Similarly, the conjunction fallacy (e.g., Tversky & Kahneman, 1981), where the joint occurrence of two events is judged more likely than either event individually, may happen because hearers pragmatically interpret the question as one of plausibility, not formal probability (Hertwig and Gigerenzer, 1999). This pragmatic motivation could conceivably be learned from training data. Previous work, however, shows that LLMs have limited inferential abilities (see, e.g., Ruis et al. (2023), where LLMs struggle to infer intention).

Different theories have been put forward on why probability matching happens. An ecological explanation draws on the tension between exploration and exploitation (e.g., Schulze et al., 2015): if 70% of the time a reward is found down path A, this path may get overrun by competitors, so it may be rational to try path B, even though it may only have a reward 30% of the time. Similarly, there is often an intuition that repeatedly choosing the most commonly correct option means missing out on learning a potentially better, or at least more interesting, strategy. Humans are pattern-seekers who tend to dislike and have difficulty correctly perceiving randomness (Bar-Hillel & Waganaar, 1991), which may make adopting a maximizing strategy (e.g., picking the above path A 100% of the time) particularly difficult. Additionally, instead of tracking probabilities and determining an optimal strategy, it appears that people sometimes follow a simple, local win-stay-lose-shift strategy (Nowak & Sigmund, 1993), where a decision is repeated if it was successful in the previous trial, otherwise a different option is chosen. Regardless of the underlying cause, probability matching often diminishes when greater cognitive resources are available (e.g., Gaissmaier & Schooler, 2008) and when a maximizing strategy is made salient (e.g., Koehler & James, 2009).

Could an LLM have learned any of these potential underlying forces from its input? Does an LLM know to behave as if it dislikes randomness, seeking out patterns where none exist? Do its actions sufficiently value exploration? Does it follow simple patterns like win-stay-lose-shift when resources are low? There are reasons to be doubtful that either optimal decisions or probability matching would be seen in LLM outputs simply because LLMs, as probabilistic language generators, perform poorly in many numeric rule-driven domains. Ravenscroft (2024), for example, prompted an LLM to respond with one term 80% of the time and with another term 20% of the time. He found, however, that that the majority term was used well above 80% of the time, and the minority term was used rarely, if ever. This boosting may be due in some part to the internal weighting that the model assigns to words (cf. ChatGPT's reported extreme preference for 42 when asked to produce a random number, vijayabhaskar96 (2024)), but it is likely also related to LLMs' inability to follow in-context deterministic rules. LLMs' performance is even more striking on simple counting tasks, such as counting the occurrences of a given word in a list of words, where performance drops dramatically as list length grows or less popular words are counted (Ball, Chen, and Herley, 2024). This highlights LLMs' dependence on statistical associations among words and their inability to perform simple deterministic, rule-based counting. When given a probability-matching task, an LLM may attempt to find patterns to follow, or it may simply pick the option that was most common in its training, or it may pick an optimal maximizing strategy but for misperceived frequencies.

2 Methods

In the current study, two repeated binary-outcome scenarios tested whether an LLM would demonstrate probability-matching behavior. The first scenario involved a coin flip and the second involved drawing a black or white ball from a bag with replacement. Each was repeated 100 times ("I have a coin whose fairness is unknown. I flipped it 100 times," and "I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times," respectively). The majority outcome (counterbalanced) was true for 70/100 trials and the minority outcome was true for the other 30/100 trials. This was presented either in summary (e.g., "70 times it came up heads, 30 times it came up tails," or "70 times I drew white, 30 times I drew black.") or as individual trials (e.g., "here are the results: TAILS, HEADS, HEADS, HEADS, …" or "here are the results: BLACK, WHITE, WHITE, WHITE, …"). GPT-4o mini was then asked to predict the next 1 or the next 10 outcomes. These prompts are illustrated in Table 1, and full prompts are provided in the Appendix.

Each prompt was presented in a new chat session and was repeated 20 times, giving 2 (Scenario) × 2 (Summary/Individual) × 2 (Majority label) × 2 (Next 1/Next 10) × 20 = 320 responses (160 1-outcome responses, 160 10-outcome responses). Prompts were submitted in November 2024, and GPT-4o mini has a reported knowledge cutoff of October 2023 (OpenAI Platform, n.d.).

Table 1. Overview of prompts, with individual outcomes truncated. The heads/tails and white/black inversions to counterbalance the majority label are not shown here (see Appendix for all stimuli).

Scenario	Intro	Summary/Individual		Next 1/Next 10	
		Summary	Individual	Next 1	Next 10
Coin	I have a coin whose fairness is unknown. I flipped it 100 times,	and 70 times it came up heads, 30 times it came up tails.	and here are the results: TAILS, HEADS, HEADS, HEADS, …	Predict the outcome of the 101st flip by responding either HEADS or TAILS.	Predict the outcome of the next 10 flips by responding either HEADS or TAILS for each flip.
Ball	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times,	and 70 times I drew white, 30 times I drew black.	and here are the results: BLACK, WHITE, WHITE, WHITE,…	Predict the outcome of the 101st draw by responding either WHITE or BLACK.	Predict the outcome of the next 10 draws by responding either WHITE or BLACK for each draw.

3 Results and Discussion

Results show overwhelmingly rational responses (i.e., the majority outcome was chosen) when predicting the Next 1 outcome, which matches behavior typically seen in humans as well. When predicting the Next 10 outcomes, GPT-4o mini appeared to engage in some degree of probability matching (Coin: 81.13% majority predictions; Ball: 84.13% majority predictions), clearly not maximizing correct predictions (100% majority predictions) but showing something like the boosting behavior seen in Ravenscroft (2024). These results, broken down by Scenario and Next 1/Next 10 are shown in Fig. 1.

For the Next-10 data, a generalized linear model shows no significant main effect of scenario (Coin, Ball) or presentation (Summarized, Individual trials), but a significant interaction between the two.

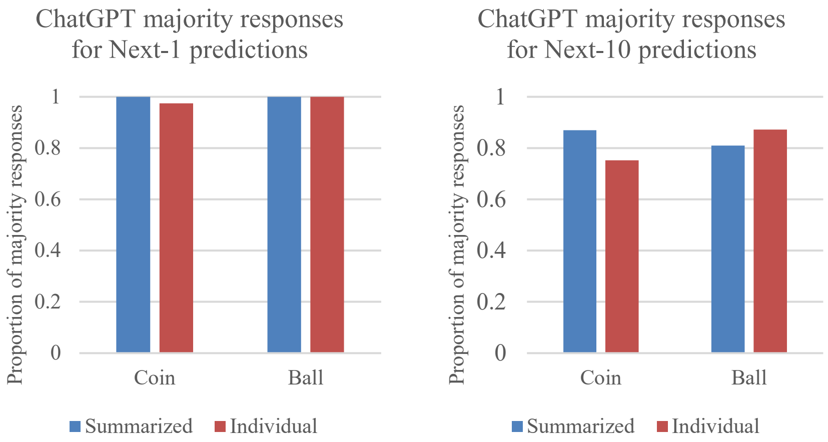

Fig. 1. Proportion of majority responses provided by ChatGPT-4o mini, shown by number of predictions (Next 1/Next 10), scenario (Coin/Ball), and presentation (Summarized/Individual).

The lack of the main effect of Scenario is somewhat surprising since coinflips in the LLM's training corpora likely tended to be fair, which could have driven the coinflip predictions closer to 50–50. There likely was not the same bias about black and white balls in the training corpora. Additionally, humans have a greater tendency to maximize when given summary information (Friedman and Massaro,1998), though it is not clear that this behavior would be well represented in the LLM's training corpora. This may also suggest that LLMs are similarly bad at probabilities (as in the summarized condition and in Ravenscroft (2024)) and counting (as in the individual trials condition and Ball et al. (2024)), such that, while probabilities and counting may be treated differently by the LLM, the results are similar.

Before providing predictions in the Individual-trials condition, GPT-4o mini usually offered a summary of the presented 100 trials either as incorrect counts (e.g., BLACK: 62; WHITE: 48) or as an incorrect description (e.g., "roughly equal"), again demonstrating LLM's inability to count. This, however, could make the above-70% majority responses more impressive and closer to an optimal solution than they may first appear. These results could represent some form of Bayesian learning, where perhaps a weaker "fairness" prior for balls versus coins allowed GPT-4o mini to move slightly closer to an optimal solution in the balls condition. Alternatively, this could be due to the same mechanism that led to boosting in Ravenscroft (2024).

4 Conclusion

This exploration in probability matching was inspired by previous work on LLMs' ability to detect patterns in longitudinal data (Zaroukian, 2024; Chandrasekaran et al., 2024; Chandrasekaran & Zaroukian, to appear). These studies found that, when asked to continue a given pattern, simple patterns were appropriately continued, but the LLM did not appear to learn more complex patterns (even reporting patterns in the data that did not exist) and instead tended to provide continuations that heavily favored certain common, often most recent, tokens from the input. In light of the work presented in the current study, there is no reason to believe the LLM in those studies was maximizing over frequencies in its input, nor was it probability matching to its input. Most likely, these are all cases of idiosyncratic model weights and limits to what associative reasoners can do with patterns, individual data points, and summarized frequencies alike. Current work explores prompt manipulations that may better situate the LLM to continue patterns, as previous studies (e.g., Mirchandani et al, 2023) have touted LLMs' strengths at in-context pattern completion tasks.

Methods to improve LLMs' abilities with similar data have been proposed, and the solution is often to outsource. For example, Nafar et al. (2024) propose improving LLM reasoning over explicit probabilities in the text (e.g., for medical decision-making) by prompting it to map probability problems to formal representations amenable to symbolic computations. Beyond this, plugins like Wolfram (2023) export certain tasks to systems capable of symbolic computation. While this may be the best solution for rational treatment of probabilities in many cases, LLMs may still hold promise in contextualizing tasks; for example, recognizing when a user is really looking for the most *relevant* answer, even when they asked for the most *probable* answer. In its current state, however, GPT-4o mini appears to perform both rationally and humanlike when giving one-off binary predictions, even if it is neither rational nor humanlike in its actual ability to count or reason over frequencies or probabilities.

Future work into the data presented here will explore the relation between the incorrect counts and the actual predictions made by GPT-4o mini, as well as the effects of other manipulations known to influence the tendency to probability match, providing a clearer picture of the shape of reasoning that LLMs provide.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Ball, T., Chen, S., Herley, C.: Can we count on LLMs? The fixed-effect fallacy and claims of GPT-4 capabilities. arXiv:2409.07638v2 (2024)
2. Bar-Hillel, M., Waganaar, W. A.: The perception of randomness. Advances in Applied Mathematics **12**(4), 428–454 (1991)
3. Chandrasekaran, A., Zaroukian E., Rawal, J., Mittrick, M., Raglin, A.: Developing a framework to evaluate credibility tracking in large language models. DEVCOM Army Research Laboratory (US), Technical Report No. ARL-TR-0057 (2024)
4. Chandrasekaran, A., Zaroukian, E.: Understanding the limitations of large language models in credibility-tracking tasks. In: Proceedings of Human-Computer Interaction (HCI) International, Springer Nature, Berlin, Germany (in process)
5. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805v2 (2018)

6. Echterhoff, J.M., Liu, Y., Alessa, A., McAuley, J., He, Z.: Cognitive bias in decision-making with LLMs. In: Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 12640–12653. Association for Computational Linguistics (2024)
7. Friedman, D., Massaro, D.W.: Understanding variability in binary and continuous choice. Psychonomic Bulletin & Review **5**(3), 370–389 (1998)
8. Gaissmaier, W., Schooler, L.J.: The smart potential behind probability matching. Cognition **109**(3), 416-422 (2008)
9. Hagendorff, T., Fabi, S., Kosinski, M.: Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. Nature Computational Science, **3**(10), 833–838 (2023)
10. Hertwig, R., Gigerenzer, G.: The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. Journal of Behavioral Decision Making **12**(4), 275–305 (1999)
11. Koehler, D., James, G.: Probability matching in choice under uncertainty: Intuition versus deliberation. Cognition **113**(1), 123-127 (2009)
12. Lin, R., Ng, H.T.: Mind the biases: quantifying cognitive biases in language model prompting in BERT. In: Findings of the Association for Computational Linguistics: ACL 2023, pp. 5269–5281. Association for Computational Linguistics, Toronto, Canada (2023)
13. Mirchandani, S., Xia, F., Florence, P., Ichter, B., Driess, D., Gonzalez Arenas, M., Rao, K., Sadigh, D., Zeng, A.: Large language models as general pattern machines. In: Proceedings of the 7th Conference on Robot Learning (CoRL). Atlanta, GA (2023)
14. Nafar, A., Venable, K.B., and Kordjamshidi, P.: Probabilistic reasoning in generative large language models. arXiv:2402.09614v1 (2024)
15. Nowak, M., Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature **364**, 56-58 (1993)
16. OpenAI Platform: Models, https://platform.openai.com/docs.models/gpt-4o, last accessed 2024/10/18
17. OpenAI: ChatGPT (GPT-3.5), https://openai.com
18. Ravenscroft, J.: LLM's can't do probability. Brainsteam. https://brainsteam.co.uk/2024/05/01/llms-cant-do-probability/, last accessed 2024/12/17
19. Ruis, L., Khan, A., Biderman, S., Hooker, S., Rocktäschel, T., Grefenstette, E.: The Gollocks of Pragmatic Understanding: Fine-Tuning Strategy Matters for Implicature Resolution by LLMs. In: Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS). Curran Associates, Inc., Red Hook, NY (2023)
20. Schultze, C., van Ravenzwaaij, D., Newell, B.R.: Of matcher and maximizers: how competition shapes choice under risk and uncertainty. Cognitive Psychology **78**, 78–98 (2015)
21. Shanks, D.R., Tunney, R.J., McCarthy, J.D.: A re-examination of probability matching and rational choice. Journal of Behavioral Decision Making, **15**(3), 233–250 (2002)
22. Suri, G., Slater, L.R., Ziaee, A., Nguyen, M.: Do large language models show decision heuristics similar to humans? A case study using GPT-3.5. arXiv:2305.04400v1 (2023)
23. Talboy, A.N., Fuller, E.: Challenging the appearance of machine intelligence: cognitive bias in LLMs and best practices for adoption. arXiv:2304.01358v3 (2023)
24. Tversky, A., Kahneman, D.: Availability: a heuristic for judging frequency and probability. Cognitive Psychology **5**(2), 207–232 (1973)
25. Tversky, A., Kahneman, D.: Judgments of and by representativeness (Report). Stanford University, Stanford, CA (1981)
26. Vijayabhaskar96. [Online forum post]. ChatGPT reflects human biases when choosing a random number but not 69. Reddit; https://www.reddit.com/r/ChatGPT/comments/1cfxt3v/chatgpt_reflects_human_biases_when_choosing_a/, last accessed 2024/12/17

27. Vulcan, N.: An economist's perspective on probability matching. Journal of Economic Surveys, **14**(1), 101–118 (2020)
28. Wolfram, S. Instant plugins for ChatGPT: introducing the Wolfram ChatGPT plugin kit. Stephen Wolfram writings. https://writings.stephenwolfram.com/2023/04/instant-plugins-for-chatgpt-introducing-the-wolfram-chatgpt-plugin-kit, last accessed 2024/12/18
29. Zaroukian, E.: Large language models for tracking reliability of information sources. In: Proceedings of Human-Computer Interaction (HCI) International, Springer Nature, Berlin, Germany (2024)

Appendix

Table A-1. All prompts used, with summary labels (not presented to the LLM) given to the left

Scenario: coin Majority label: heads Presentation: summary Predictions: next 1	I have a coin whose fairness is unknown. I flipped it 100 times, and 70 times it came up heads, 30 times it came up tails. Predict the outcome of the 101st flip by responding either HEADS or TAILS.
Scenario: coin Majority label: tails Presentation: summary Predictions: next 1	I have a coin whose fairness is unknown. I flipped it 100 times, and 30 times it came up heads, 70 times it came up tails. Predict the outcome of the 101st flip by responding either HEADS or TAILS.
Scenario: coin Majority label: heads Presentation: summary Predictions: next 10	I have a coin whose fairness is unknown. I flipped it 100 times, and 70 times it came up heads, 30 times it came up tails. Predict the outcome of the next 10 flips by responding either HEADS or TAILS for each flip.
Scenario: coin Majority label: tails Presentation: summary Predictions: next 10	I have a coin whose fairness is unknown. I flipped it 100 times, and 30 times it came up heads, 70 times it came up tails. Predict the outcome of the next 10 flips by responding either HEADS or TAILS for each flip.
Scenario: coin Majority label: heads Presentation: individual Predictions: next 1	I have a coin whose fairness is unknown. I flipped it 100 times, and here are the results: TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, HEADS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, HEADS, TAILS, HEADS, TAILS. Predict the outcome of the 101st flip by responding either HEADS or TAILS.

Scenario: coin Majority label: tails Presentation: individual Predictions: next 1	I have a coin whose fairness is unknown. I flipped it 100 times, and here are the results: TAILS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, HEADS, TAILS, TAILS, HEADS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS, HEADS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS. Predict the outcome of the 101st flip by responding either HEADS or TAILS.
Scenario: coin Majority label: heads Presentation: individual Predictions: next 10	I have a coin whose fairness is unknown. I flipped it 100 times, and here are the results: TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, HEADS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, HEADS, HEADS, HEADS, TAILS, HEADS, TAILS. Predict the outcome of the next 10 flips responding either HEADS or TAILS for each flip.
Scenario: coin Majority label: tails Presentation: individual Predictions: next 10	I have a coin whose fairness is unknown. I flipped it 100 times, and here are the results: TAILS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, HEADS, TAILS, TAILS, HEADS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, HEADS, TAILS, HEADS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, HEADS, HEADS, HEADS, TAILS, TAILS, TAILS, HEADS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS, TAILS,

	TAILS, HEADS, TAILS, HEADS, TAILS, TAILS, TAILS, TAILS, HEADS, TAILS. Predict the outcome of the next 10 flips responding either HEADS or TAILS for each flip.
Scenario: ball Majority label: white Presentation: summary Predictions: next 1	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and 70 times I drew white, 30 times I drew black. Predict the outcome of the 101st draw by responding either WHITE or BLACK.
Scenario: ball Majority label: black Presentation: summary Predictions: next 1	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and 30 times I drew white, 70 times I drew black. Predict the outcome of the 101st draw by responding either WHITE or BLACK.
Scenario: ball Majority label: white Presentation: summary Predictions: next 10	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and 70 times I drew white, 30 times I drew black. Predict the outcome of the next 10 draws by responding either WHITE or BLACK for each draw.
Scenario: ball Majority label: black Presentation: summary Predictions: next 10	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and 30 times I drew white, 70 times I drew black. Predict the outcome of the next 10 draws by responding either WHITE or BLACK for each draw.
Scenario: ball Majority label: white Presentation: individual Predictions: next 1	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and here are the results: BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, WHITE, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, WHITE, BLACK, WHITE, BLACK. Predict the outcome of the 101st draw by responding either WHITE or BLACK.
Scenario: ball Majority label: black Presentation: individual Predictions: next 1	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and here are the results: BLACK, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, WHITE, BLACK, BLACK, WHITE, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK,

	BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK, WHITE, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK. Predict the outcome of the 101st draw by responding either WHITE or BLACK.
Scenario: ball Majority label: white Presentation: individual Predictions: next 10	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and here are the results: BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, WHITE, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, WHITE, WHITE, WHITE, BLACK, WHITE, BLACK. Predict the outcome of the next 10 draws by responding either WHITE or BLACK for each draw.
Scenario: ball Majority label: black Presentation: individual Predictions: next 10	I have a bag of an unknown number of black balls and white balls. I drew one ball with replacement 100 times, and here are the results: BLACK, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, WHITE, BLACK, BLACK, WHITE, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, WHITE, BLACK, WHITE, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, WHITE, WHITE, WHITE, BLACK, BLACK, BLACK, WHITE, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK, WHITE, BLACK, BLACK, BLACK, BLACK, WHITE, BLACK. Predict the outcome of the next 10 draws by responding either WHITE or BLACK for each draw.