The Illusion of Randomness: How LLMs Fail to Emulate Stochastic Decision-Making in Rock-Paper-Scissors Games?

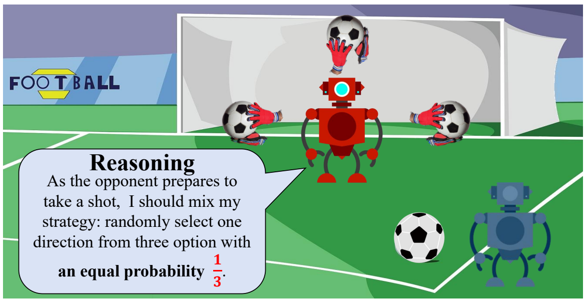
Zihao Guo[1], **Hongtao Lv**[1*], **Chaoli Zhang**[2], **Yibowen Zhao**[1], **Yixin Zhang**[1], **Lizhen Cui**[1*]
[1] School of Software & Joint SDU-NTU Centre for Artificial Intelligence Research,
Shandong University, China
[2] School of Computer Science and Technology, Zhejiang Normal University, China
`{zihaog, ybw.zhao, yixinzhang}@mail.sdu.edu.cn,`
`{lht, clz}@sdu.edu.cn, chaolizcl@zjnu.edu.cn`

Abstract

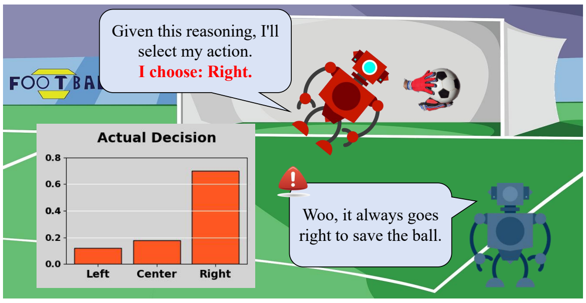
Prior research indicates that although large language models (LLMs) can precisely articulate the theoretical probability distributions associated with optimal strategic choices, their actual decision-making systematically diverges from these prescriptions—a phenomenon we define as the cognition–behaviour gap in LLMs. For example, in a Rock–Paper–Scissors (RPS) game, LLMs correctly identify the strategy of Nash equilibrium as selecting each action (Rock, Paper, Scissors) with equal probability $\frac{1}{3}$, but their observed choices systematically deviate from this uniform distribution. Through a comprehensive evaluation of 20 state-of-the-art LLMs, we identify two critical insights: (1) we demonstrate that intrinsic biases inherited from pre-training corpora alone are insufficient to explain the observed deviations; (2) we introduce a semantic-free paradigm that strips away intrinsic biases to isolate pure positional bias-LLMs exhibit distinct position preferences—for example, o1 favours the first option, DeepSeek-V3 peaks the middle and DeepSeek-R1 shows a bimodal bias toward first and last positions. Our findings advocate innovation to bridge the gap between strategic reasoning and decision-making in LLMs.

1 Introduction

Large language models (LLMs) have demonstrated remarkable capabilities in strategic reasoning tasks, from solving mathematical games (Ahn et al., 2024) to simulating negotiations (Bianchi et al., 2024). However, existing research shows that the stochastic decision-making of LLMs remains deficient (Van Koevering and Kleinberg, 2024), and LLMs deviate from ideal randomness in tasks as simple as coin-flipping (Van Koevering and Kleinberg, 2024), dice-rolling (Liu, 2023), and pseudo-random number generation (Hopkins and Renda, 2023). In particular, such deviations persist when LLMs explicitly attempt to approximate mixed-strategy Nash equilibrium (Silva, 2024; Poje et al., 2024). For instance, in the Rock–Paper–Scissors (RPS) game, (Xu et al., 2024) reports that GPT-4 correctly identifies the Nash equilibrium as uniform randomization (33.3% per action) yet selects Rock in 67% of 100 independent trials—approximately twice the theoretical frequency.


FOOTBALL
Reasoning
As the opponent prepares to take a shot, I should mix my strategy: randomly select one direction from three option with an equal probability $\frac{1}{3}$.

(a) Reasoning


FOOTBALL
Given this reasoning, I'll select my action. I choose: Right.
Actual Decision
0.8
0.6
0.4
0.2
0.0
Left
Center
Right
Woo, it always goes right to save the ball.

(b) Behaviour

Figure 1: A toy example in an embodied-intelligence setting: two LLM-driven robots face off in a penalty shootout.

This limitation may pose an obstacle to real-world deployments, for instance, LLM-driven agents in embodied intelligence. Consider a simplified, illustrative scenario involving two robots engaged in a penalty shootout, as depicted in Figure 1. In this hypothetical situation, an LLM-controlled goalkeeper correctly identifies the Nash equilibrium strategy, which is diving left, center, or right

---

[*] Corresponding authors.

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 8618–8637
November 4-9, 2025 ©2025 Association for Computational Linguistics

with equal probability. But if its actual decisions skew, say, 70% to the right, an opposing kicker can learn and exploit that bias, dramatically undermining the goalkeeper's effectiveness. We call this mismatch between what the LLMs know and what they do as the "cognition–behaviour gap".

Initially, Meister et al. (2024) attributes such gaps to failures in probabilistic sampling despite correct distributional reasoning. A prevailing explanation further links these failures to intrinsic biases inherited from pre-training corpora (Xu et al., 2024; Guo et al., 2024). However, the evidence remains tentative: existing results cover only a narrow set of models and experimental designs, and no systematic, cross-model analysis has cleanly disentangled corpus-level effects from other confounders, which motivates the present investigation.

In this work, we center our analysis on RPS, whose unique mixed-strategy Nash equilibrium prescribes uniform random play. We comprehensively evaluate 20 state-of-the-art LLMs to examine how LLMs fail to emulate this stochastic decision-making. We find that nearly all advanced LLMs (e.g., GPT-4.5 (OpenAI, 2025a), DeepSeek-R1 (Guo et al., 2025)) exhibit systematic biases, with empirical action distributions deviating from the probability distributions they explicitly derive. We establish two insightful findings:

(1) Intrinsic biases in pre-training corpora alone fail to explain the gap. We first examine whether LLM choice reflects token frequencies in their training data, but we found a clear mismatch-most LLMs over-selected `rock` despite `paper` being most common in corpora. Thus, we introduce a semantic-free prompting paradigm by replacing each action with a random 10-character string to remove all lexical-frequency and semantic cues. Empirical results show that LLMs continue to exhibit pronounced choice skews, confirming that corpus-level intrinsic biases alone cannot explain the cognition–behaviour gap in LLM stochastic decision-making.

(2) Positional bias exerts a significant influence. By removing the semantic bias (Lin and Li, 2025) and lexical frequency bias (Zhao et al., 2021) introduced by the tokens, we are the first to isolate pure positional bias in stochastic decision-making. Under these conditions, distinct LLMs display characteristic position preferences—for example, o1 favours the first option, DeepSeek-V3 peaks on middle entries, and DeepSeek-R1 shows a bimodal bias toward the initial and last positions.

The remainder of this paper is structured as follows: Section 2 synthesises related work on the gap of randomness behaviour in LLMs. Section 3 introduces the background of our research. Section 4 presents empirical results to declare our findings. Section 5 discusses the positional-bias failure in stochastic choices and outlines practical mitigations. Section 6 concludes our empirical findings.

2 Related Work

Recent work has highlighted significant challenges in LLMs' abilities to generate truly random samples. For instance, when GPT-4 "rolls" a virtual die, it outputs some faces far more often than the theoretical $\frac{1}{6}$ frequency. In longer sequences, the marginal distribution becomes nearly uniform, yet adjacent numbers repeat less often than chance would predict (Liu, 2023). Even the ostensibly simpler task of producing binary sequences is affected: GPT-4 and LLaMA 3 reproduce human-like cognitive biases in simulated coin flips, whereas GPT-3.5 exhibits behaviour that more closely approximates randomness (Van Koevering and Kleinberg, 2024). These errors persist even when models are asked to sample from arbitrary target distributions (Hopkins and Renda, 2023), suggesting that the difficulty is agnostic to output space. The gap is also evident in game-theoretic settings. In RPS, for example, Xu et al. (2024) show that GPT-4 correctly states the uniform mixed-strategy Nash equilibrium yet still chooses `rock` in 67% of independent rounds—roughly twice the prescribed 33% frequency. A similar mismatch has been documented in other game scenarios like Matching Pennies, Chicken Games, etc. across GPT-3.5 and GPT-4 (Silva, 2024; Poje et al., 2024), indicating a consistent failure to translate theoretical distributions into action.

Xu et al. (2024) attributes the cognition-behaviour gap, in which the mismatch between LLMs' stated understanding of probability and their generated outputs to intrinsic bias inherited from the pre-training corpora. Several studies document a "frequency effect" as the factor of intrinsic bias, where token prevalence in pre-training corpora systematically skews LLM outputs (Lovering et al., 2024; Wei et al., 2021; McCoy et al., 2024). Building upon this claim, we want to uncover how more frequent words can dominate stochastic generation even when uniform randomness is prescribed. Recent work has brought attention to positional bias—the tendency of LLMs to favour answer options based on their placement within the prompt (Pezeshkpour and Hruschka, 2023). For example, Lovering et al. (2024) show that when LLMs are asked to choose between two colours, GPT-4o-mini consistently selects the first option, while LLaMA-3.1-8B favours the second. However, such colour preference questions inherently engage corpus-driven biases about common colour associations. To further study it, we disentangle the source of biases and accurately measure the impact of positional bias on LLM decision-making.

Prior works have identified and mitigated LLM selection biases in multiple-choice question (MCQ) tasks (Pezeshkpour and Hruschka, 2023). For example, (Wei et al., 2024) proposed grey-box probability calibration and a black-box two-hop querying strategy to counter option-order biases. (Choi et al., 2024) further introduced a black-box Bias Node Pruning method to remove internal bias parameters, as well as an Auxiliary Option Injection technique that adds a dummy "I don't know" choice to flatten the selection distribution. However, these approaches assume an objectively correct answer and often require numerous model queries or direct model access, making them less applicable to our scenarios, where all options are equally valid (e.g., RPS) and efficiency is paramount.

3 Background

3.1 Rock-Paper-Scissors

Rock–Paper–Scissors (RPS)—also known as Rochambeau or Jan-Ken-Pon—is a canonical game whose unique solution is a mixed-strategy Nash equilibrium. In RPS, each participant's strategy space in RPS is given by:

$$S = \{\mathcal{R}, \mathcal{P}, \mathcal{S}\}, \tag{1}$$

where $\mathcal{R}$ represents action option Rock, $\mathcal{P}$ represents Paper, $\mathcal{S}$ represents Scissors. The game rules establish a cyclic dominance: $\mathcal{R}$ beats $\mathcal{S}$, $\mathcal{S}$ cuts $\mathcal{P}$, and $\mathcal{P}$ wraps $\mathcal{R}$; identical actions result in a tie. Each player has a probability P_i to choose an action i, where $i \in \{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$.

After confirming that every LLM can identify the mixed-strategy Nash equilibrium of RPS (i.e., a uniform $P_i = \frac{1}{3}$ distribution over $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$), we evaluate whether they can implement it. For each LLM, we run T independent rounds ($T = 100$ in our experiments). In every round t, LLM is required to commit a single action $A_t \in \{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$. We define the empirical selection frequencies:

$$\hat{P}_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{A_t = i\}, \qquad i \in \{\mathcal{R}, \mathcal{P}, \mathcal{S}\}. \tag{2}$$

We then compare the vector $\hat{\boldsymbol{P}} = (\hat{P}_{\mathcal{R}}, \hat{P}_{\mathcal{P}}, \hat{P}_{\mathcal{S}})$ with the ideal uniform vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ in the game round t though a probabilistic analysis framework. The resulting divergence scores measure how far each LLM's decision departs from equilibrium-consistent, truly random play.

3.2 Evaluation of Stochastic Processes

Inspired by (Gupta et al., 2025), we adopt an evaluation of stochastic processes framework that quantifies the divergence between the theoretical uniform distribution $\frac{1}{3}$–$\frac{1}{3}$–$\frac{1}{3}$ in RPS and the empirical choice distribution produced by LLMs. Given a fixed prompt, $P_M(A \mid \text{prompt})$ represents the conditional probability of action A estimated by the LLM M in a 100-trial sampling procedure. To handle potential invalid actions outside of S, the linear normalization approach is used:

$$\hat{P}_M(A) = \frac{P_M(A \mid \text{prompt})}{\sum_{A' \in S} P_M(A' \mid \text{prompt})}. \tag{3}$$

To quantify the discrepancy between the normalized empirical distribution $\hat{P}_M(A)$ and the target distribution $P^*(A)$ prescribed under rational play, we compute the Total Variation Distance (TVD):

$$\delta\Big(P^*, \hat{P}_M\Big) = \frac{1}{2} \sum_{A \in S} \big|P^*(A) - \hat{P}_M(A)\big|. \tag{4}$$

where $\delta\Big(P^*, \hat{P}_M\Big) \in [0, 1]$; a value of 0 denotes perfect alignment, whereas larger values indicate greater divergence from the expected distribution. In the context of RPS, the strategic distribution is $P^*(A) = \frac{1}{3}, \forall A \in \{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$, and hence, $\delta\Big(P^*, \hat{P}_M\Big) \in [0, \frac{2}{3}]$.

4 Experiment

We begin with a case study to visualize action distributions of 20 LLMs to confirm the presence of a cognition–behaviour gap. Then, we examine its underlying determinants by systematically analysing two factors: (1) intrinsic bias inherited from pre-training corpora, and (2) positional bias arising from the ordering of answer options within prompts. More experiments, discussions, and explanations can be found in Appendix A.

4.1 Experiment Setup

Models. The evaluated architectures encompass canonical and Long Chian-of-Thought (CoT) reasoning LLMs. Selected models include GPT-4.5-preview-2025-02-27 (OpenAI, 2025a), o3-mini (OpenAI, 2025b), o1-2024-12-17 (Jaech et al., 2024), o1-mini-2024-09-12, GPT-4o-2024-08-06 (Hurst et al., 2024), GPT-4-turbo (Achiam et al., 2023), GPT-3.5-turbo (Ye et al., 2023), Claude-3-7-sonnet-20250219 (Anthropic, 2025), Claude-3-5-sonnet-20240620 (Anthropic, 2024b), claude-3-haiku-20240307 (Anthropic, 2024a), Gemini-2.0-pro-exp-02-05 (Google DeepMind, 2025), Gemini-2.0-flash, Gemini-1.5-pro-latest (Team et al., 2024), Gemini-1.5-flash-latest, Meta-LLaMA-3.1-8B-Instruct (Grattafiori et al., 2024), Meta-LLaMA-3.1-405B-Instruct, DeepSeek-V3-250324 (Liu et al., 2024), DeepSeek-R1-250120 (Guo et al., 2025), Grok-3 (xAI, 2025), and Qwen-2.5-max (Team, 2024).

Configurations. All model evaluations were conducted in a vanilla state using the official API interface[1]. To ensure statistical significance, each experimental scenario was tested on 100 independent games, with a new session established for each call to eliminate any contextual interference. We conduct an initial experiment involving 1,000 independent generations from the LLMs in Appendix A.1. The results show that 100 runs are sufficient to approximate the model's output distribution. The experimental parameters were fixed at temperature = 1, top-p = 1, and repetition penalties were disabled, ensuring nondeterministic generation and consistent experimental settings.

4.2 Cognition-Behaviour Gap

In this section, we corroborate the finding of (Meister et al., 2024) that LLMs are better at describing probability distributions than at sampling from them. we tested the individual LLMs using the same Prompt[2] provided by (Xu et al., 2024).

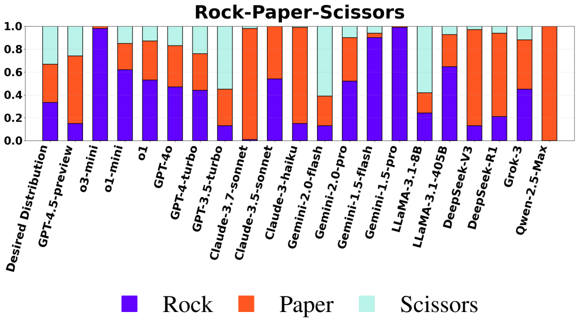


Figure 2: The choice distribution of LLMs when playing Rock-Paper-Scissors. The x-axis lists the evaluated LLM models, while the y-axis reports the share of `rock`, `paper`, and `scissors` selections that normalized over 100 independent games.

Models	TVD (↓)	Models	TVD (↓)
GPT-4.5-preview	0.257	Gemini-2.0-flash	0.277
o3-mini	0.647	Gemini-2.0-pro	0.233
o1-mini	0.287	Gemini-1.5-flash	0.567
o1	0.203	Gemini-1.5-pro	0.657
GPT-4o	0.163	LLaMA-3.1-8B	0.247
GPT-4-turbo	0.107	LLaMA-3.1-405B	0.313
GPT-3.5	0.217	DeepSeek-V3	0.507
Claude-3.7	0.637	DeepSeek-R1	0.397
Claude-3.5	0.333	Grok-3	0.213
Claude-3	0.537	Qwen-2.5-Max	0.667

Table 1: The mean total variation distance (TVD, ↓) across tested LLMs, larger values indicate greater divergence from the expected probability $\frac{1}{3}$.

> **Prompt**: You are playing the RPS game. You should first reason about the Nash equilibrium of this game, and then choose one action from $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ based on your reasoning. Please choose an exact action.

In the experimental results, all tested LLMs correctly state that the unique equilibrium in RPS is to choose $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ with equal probability $\frac{1}{3}$, an example is provided in Appendix A.2. We then compare this uniform benchmark with the action choice frequencies over 100 independent rounds. Figure 2 plots the empirical distributions, while Table 1 reports the numerical results that record the mean total variation distance, δ, between each model's output and the theoretical baseline. It is striking to find that several models like GPT-o3-mini, Claude-3.7, and Grok-3 exhibit extreme skew ($\delta > 0.6$), effectively locking onto a single action. These quantitative results substantiate the gap in stochastic decision-making.

---

[1] All API accesses strictly followed the license agreements and terms of service of respective API providers (OpenAI, Anthropic, Google, Meta, etc.), used solely for academic research purposes without commercial benefit. Our experimental protocol complied with all prescribed usage limitations, including rate limits, output restrictions, and content policies outlined in each provider's academic access terms.

[2] Noted that RPS in real Prompt is expressed as `Rock-Paper-Scissors`, and $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ are `Rock`, `Paper`, `Scissors`, respectively.

4.3 Intrinsic Bias

Recent work has documented a frequency effect, whereby token frequency in pre-training corpora systematically biases LLM outputs (Lovering et al., 2024; Wei et al., 2021; McCoy et al., 2024). We therefore hypothesize that LLMs' stochastic decisions will reflect these frequency priors: actions labelled by higher-frequency tokens should be chosen more often. We expect LLMs to disproportionately favour actions whose labels appear most frequently in their training data. Since the pre-training corpora of closed-source LLMs are inaccessible, we estimate their token-frequency priors using large public datasets as proxies. Prior work shows that GPT-4o and Gemini 1.5 Pro replicate the lexical frequency ordering of the Corpus of Contemporary American English (COCA) with near-perfect fidelity (Davies, 2025), validating this proxy approach for assessing the relative frequencies of `rock`, `paper`, and `scissors`.

Building on this insight, we analyze the occurrence frequency of `rock`, `paper`, and `scissors` in open-source corpora hosted on *English-Corpora.org*[3], *Google Books n-grams*[4], *Colossal Clean Crawled Corpus (C4)* (Raffel et al., 2020)[5], *RedPajama* (Weber et al., 2024) [6], *Dolma* (Soldaini et al., 2024)[7]. The result places `paper` first, `rock` second, and `scissors` third.

If the occurrence frequency of `rock`, `paper`, and `scissors` constitutes the main reason for stochastic decision-making of LLMs, then they should over-produce $\mathcal{P}$. However, our experiments reveal that the majority of LLMs disproportionately select $\mathcal{R}$, even though `paper` is the most frequent token in reference corpora. This finding indicates that token frequency alone cannot account for the observed bias. We therefore turn to a second form of intrinsic bias: semantic salience arising from human cognitive associations. For example, `rock` is commonly linked to strength metaphors in language (Zhang et al., 2021) and exhibits a mild selection advantage among human players in RPS games (Dyson et al., 2016). Such semantic priors may drive the persistent `rock` preference in LLM decision-making.

To isolate non-lexical factors, we introduce a semantic-free paradigm in which each action label is replaced by a randomly generated, non-pronounceable 10-character string (i.e., `orxtlwsjuf, qrfnrzsosi, bvmfdwqdes`). The rest of the prompt, including the RPS rules and the prescription of the mixed-strategy Nash equilibrium, remains unchanged (see Appendix A.2). As Figure 3 shows, the models' choices remain far from the uniform distribution. This finding indicates that, although intrinsic biases—whether due to word frequency or semantic associations—undeniably shape LLM choices, they alone are insufficient to explain the persistent divergence from the uniform mixed-strategy equilibrium.

Corpus Name	**rock**	**paper**	**scissors**
English Corpora			
COCA	87,552	**126,476**	4,195
COHA	41,262	**73,508**	2,206
Google Books (American English)	7,000,769	**19,070,293**	511,265
Google Books (British)	2,307,015	**5,225,572**	120,782
n-gram Language Models			
C4	10,670,341	**23,690,184**	492,327
RedPajama	72,721,463	**106,761,993**	1,406,924
Dolma	113,672,854	**274,465,797**	4,943,023

Table 2: Open-source corpora.

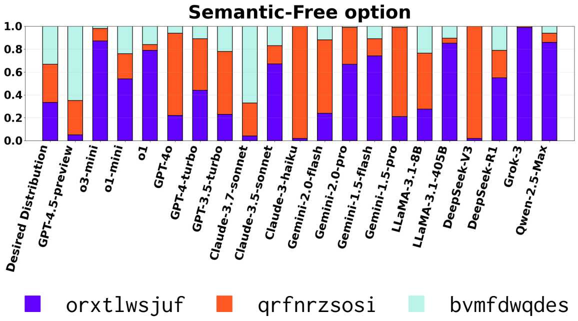


Figure 3: The choice distributions of LLMs in a semantic-free variant of Rock–Paper–Scissors game, where a randomly-generated, semantic-free string replaces each action label.

4.4 Positional Bias

Inspired by Pezeshkpour and Hruschka (2023), which demonstrates that LLMs are highly sensitive to option order and, when uncertain, tend to favour specific positions. However, their experiments are confined to reasoning tasks rather than stochastic decision-making. We therefore extend the investigation to a purely stochastic setting, testing whether positional bias similarly distorts LLM decision-making on randomness.

---

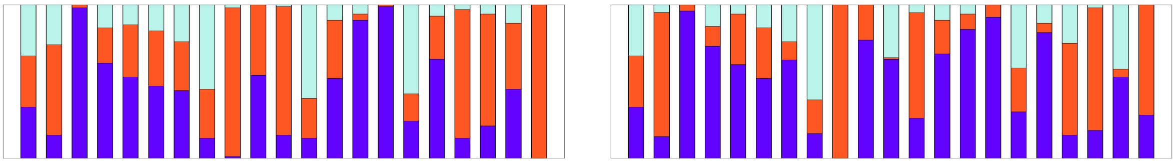
[3] `https://www.english-corpora.org//`

[4] `https://books.google.com/ngrams/`

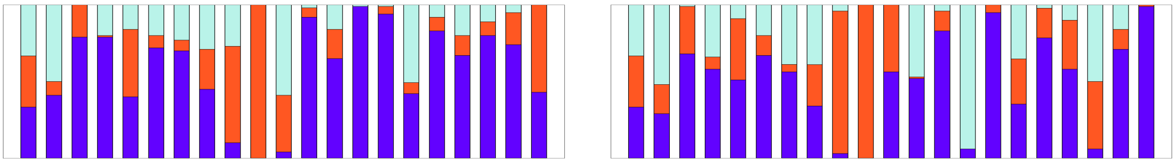
[5] `https://huggingface.co/datasets/allenai/c4`

[6] `https://huggingface.co/datasets/togethercomputer/RedPajama-Data-1T`

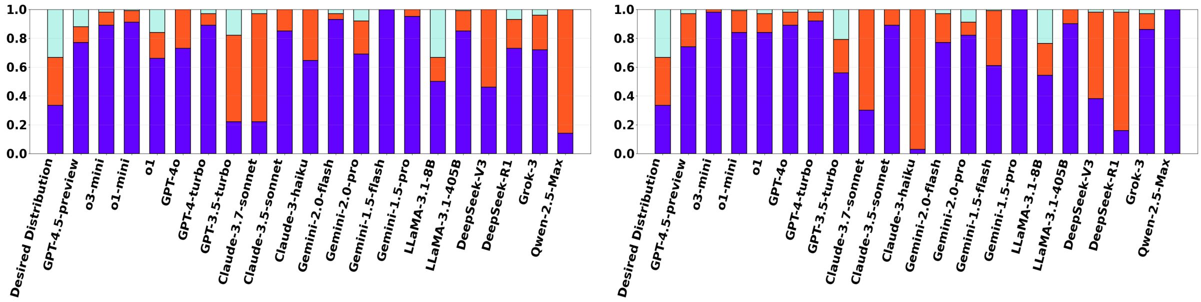
[7] `https://huggingface.co/datasets/allenai/dolma`



(a) **Rock-Paper-Scissors** (b) **Rock-Scissors-Paper**



(c) **Paper-Scissors-Rock** (d) **Paper-Rock-Scissors**



(e) **Scissors-Paper-Rock** (f) **Scissors-Rock-Paper**

🟪 Rock 🟧 Paper 🟦 Scissors

Figure 4: **The choice distribution variations of LLMs across Rock-Paper-Scissors Prompt order permutations.** Six lexical permutations were used to describe the game in the prompt. The x-axis lists the evaluated LLM models, while the y-axis reports the share of `rock`, `paper`, and `scissors` selections that normalized over 100 independent games. **Panels (a)–(d) use the same x-axis and y-axis labels as panels (e) and (f).**

4.4.1 Rock-Paper-Scissors

Firstly, we extend experiments on RPS games to examine the existence of positional bias. We conduct a fully crossed model experiment incorporating all six permutations of the $\mathcal{R}/\mathcal{P}/\mathcal{S}$ triad. The canonical $\mathcal{R}$-$\mathcal{P}$-$\mathcal{S}$ sequence serves as the control condition, while the remaining five permutations constitute the treatment set. Every prompt follows a fixed template:

> **Prompt**: You are playing the [RPS/RSP/PSR /PRS/SPR/SRP] game. You should first reason about the Nash equilibrium of this game, and then choose one action from [$\{\mathcal{R},\mathcal{P},\mathcal{S}\}$/ $\{\mathcal{R},\mathcal{S},\mathcal{P}\}$/$\{\mathcal{P},\mathcal{S},\mathcal{R}\}$/$\{\mathcal{P},\mathcal{R},\mathcal{S}\}$/$\{\mathcal{S},\mathcal{P},\mathcal{R}\}$/ $\{\mathcal{S},\mathcal{R},\mathcal{P}\}$] based on your reasoning. Please choose an exact action.

As a result, we find that even if we change lexical permutations, all the tested LLMs can still recognize the mixed-strategy Nash equilibrium and believe that the rational choice should be to select each action with a probability of $\frac{1}{3}$. Figure 4 shows that the empirical action frequencies of most LLMs shift across the six prompt permutations, confirming that option order—the hallmark of positional bias—influences their decisions. Diverse decision patterns emerged among several models:

Rock preference:11 of the 20 evaluated LLMs consistently favour $\mathcal{R}$ across all six permutations of option order, indicating a stable bias that is unaffected by re-ordering.

Median preference: Qwen 2.5 Max consistently selects the middle-positioned action, choosing $\mathcal{R}$ or $\mathcal{P}$ near 100% frequency when centrally placed, but exhibiting a bimodal distribution when $\mathcal{S}$ occupies the middle position.

Counter pattern: Interestingly, GPT-4.5-preview systematically chooses the action that defeats the first item in the prompt sequence. When the list starts with $\mathcal{R}$, the model selects in $\mathcal{P} \approx 70\% \pm 10\%$ of trials; when it starts with $\mathcal{P}$, it shifts to $\mathcal{S}$ ($\approx 50\% \pm 1\%$); and when it starts with $\mathcal{S}$, it moves to $\mathcal{R}$ ($\approx 75\% \pm 2\%$). Qualitative inspection of the model's response suggests that it im-

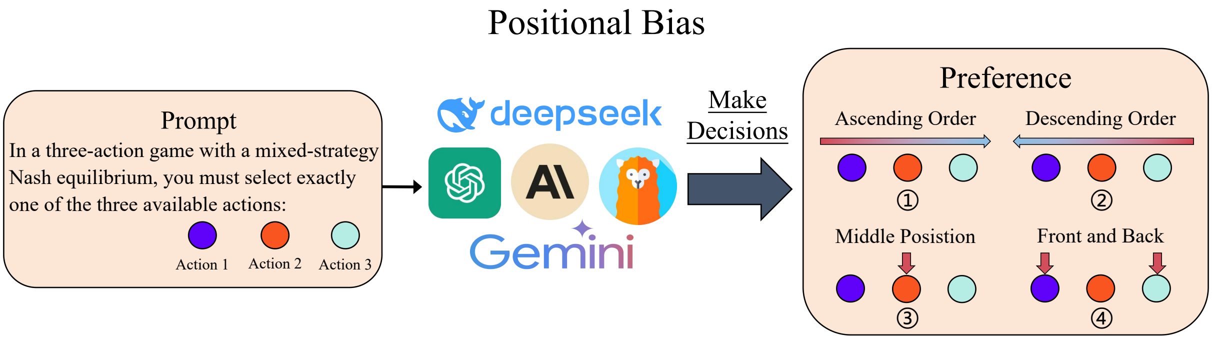
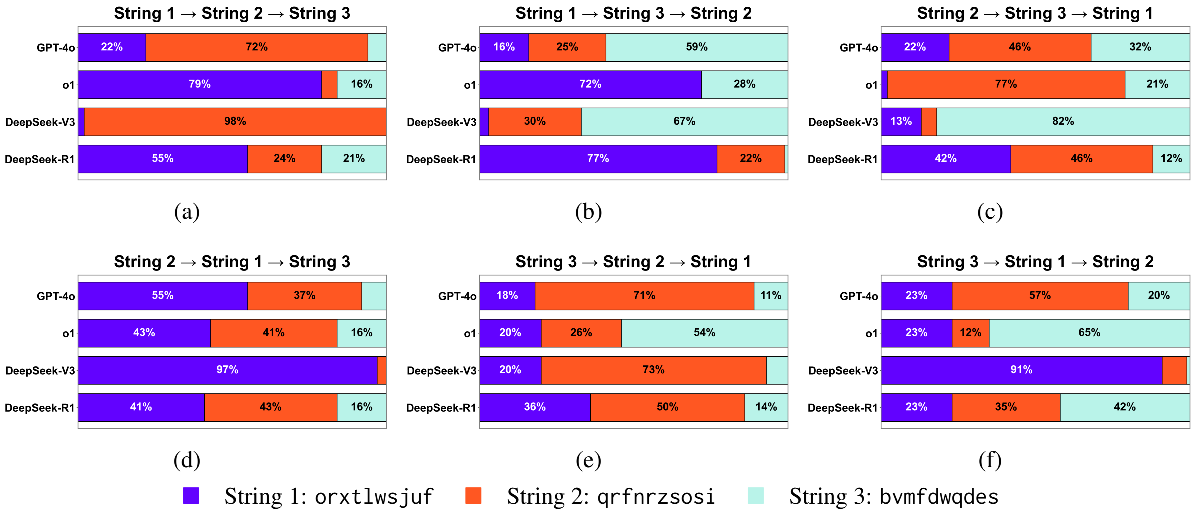


Figure 5: The Positional Bias of Prompt may interfere with decisions of LLMs in mixed-strategy Nash equilibrium.



(a) (b) (c)

(d) (e) (f)

Figure 6: **The six lexical permutations used to describe the game illustrate the positional bias on LLMs' decision-making.** Each subfigure reports the normalised choice distribution across LLMs. The replaced element for `rock`, `paper`, `scissors` are `orxtlwsjuf`, `qrfnrzsosi`, or `bvmfdwqdes`, respectively.

plicitly treats the first-listed option as the opponent's likely move and therefore responds with the counter-action that would win.

In this experiment, we retained intrinsic bias, which continues to skew LLMs toward $\mathcal{R}$—but permuted the positions of the answer options. The resulting shifts in choice distributions reveal the additional influence of positional bias.

4.4.2 Semantic-Free Paradigm

To further disentangle the effect of positional bias in Prompt from the confounding influence of intrinsic biases in the corpus, we continue to implement the semantic-free paradigm to illustrate this effect. Under this controlled setting, we permute the order of the three actions and compare the resulting choice distributions across models. Because of page-length constraints, this part of the experiments is restricted to four models: GPT-4o, o1, DeepSeek-V3, and DeepSeek-R1.

As depicted in Figure 6, our evaluation across six permutation conditions reveals that none of the tested LLMs achieved a uniform distribution. Instead, three distinct position-dependent patterns emerged as introduced in the previous context: (1) o1 and DeepSeek-R1 exhibited ascending-order preference, selecting the first-position option with 55%-77% probability when "String 1" led the sequence, and diminished to 2%-36% when "String 1" moved to terminal positions, demonstrating position-index correlated sensitivity; (2) GPT-4o and DeepSeek-V3 consistently preferred median-position options across all valid permutations; This positional bias dominance persists even when intrinsic biases are eliminated, confirming spatial encoding's critical role in LLM decision-making. Moreover, Long CoT reasoning models (o1 and DeepSeek-R1) consistently over-select the first option in the sequence, whereas GPT-4o and DeepSeek-V3 reliably favour the middle option. However, this three-option game offers limited evidence of positional bias. To bolster our findings,

Length	Sequence	Models (%)			
		4o	o1	V3	R1
4	nalmozkzhf	41	70	14	73
	qlqfwlklnw	33	6	4	0
	riwpharsgy	26	15	82	19
	vfjhmayacd	0	9	0	8
5	jjhgelerzs	13	53	9	57
	hcqefmxrac	12	6	0	0
	ulqgtnwpwg	66	22	90	2
	yfhodxulzn	7	19	1	0
	dozxichhrn	2	0	0	41
6	oxaatanhzh	18	48	20	60
	nrpjvitdeb	8	2	0	1
	jmakgkoepx	24	17	77	4
	zgpkjexdwt	41	23	3	3
	vyanqddyfn	7	0	0	0
	jkaawzkgya	2	10	0	32
7	crrzcdsfrk	20	71	27	73
	zsydssgddt	7	1	0	0
	oeoaiylxad	31	7	12	0
	xgranflukp	15	20	45	5
	vcwoazslgz	14	1	0	0
	cmorbvkjzi	8	0	3	11
	jmllurvola	5	0	13	11

Table 3: Results of the extended-sequence experiment: for each LLM, the percentage frequency with which options are selected across 4-, 5-, 6-, and 7-option games.

we extend our experiments to games with larger action sets to examine whether this bias persists or changes as the choice space grows.

4.4.3 Multi-option Games

To clarify whether positional bias in LLM decisions arises from absolute positions or relative ordering, we extend our experiments to n-option sequential games ($n = 4, 5, 6, 7$) with unique uniform mixed-strategy Nash equilibrium solutions, where the theoretical solution of each action evolves as $\frac{1}{n}$. The experimental prompt builds upon the previous design outlined in Appendix A.2, with the modification of the cycle patterns' length. As shown in Table 3, once the prompt contains more than four options ($n > 4$), all models largely discard their earlier ordinal preferences, and all models tend to abandon ordinal preferences. Specifically, GPT-4o and DeepSeek-V3 redistribute decision weights toward the first-position selection while maintaining primary preference on median indices. Further, we find that o1 exhibits stabilized first-position selection probability and secondarily prefers third and fourth elements, whereas DeepSeek-R1 develops a bimodal distribution peaking at terminal positions of the sequence. These findings demonstrate that, in longer sequences, position preference rather than relative order drives LLM action selection in uniform play, thereby highlighting positional rather than ordinal mechanisms as the core contributor to stochastic decision-making.

4.5 Support Experiments

The supplementary material provides additional experimental blocks that extend our findings.

(1) **Model scale does not mitigate randomness deficits.** In Appendix A.3, we compare LLMs of increasing model scale within the same architecture family and find that larger scales do not reduce the cognition–behaviour gap in RPS. Thus, simply scaling up an LLM is insufficient to eliminate its shortcomings in stochastic decision-making.

(2) **Generalization to other mixed-strategy games.** In Appendix A.4, we extend our evaluation to two additional mixed-strategy games—Matching Pennies and Morra. The cognition–behaviour gap persists in both cases, confirming that the randomness shortfall is not peculiar to RPS but generalises across diverse strategic settings

(3) **Temperature tuning.** In Appendix A.5, we vary the LLM sampling temperature—from low (deterministic) to high (high-entropy) settings— to examine whether LLMs' decision is influenced by output entropy. Our results indicate that the cognition–behaviour gap persists across the temperature range, indicating that observed biases cannot be attributed to the entropy of the sampling step.

(4) **Prompt-design sensitivity.** In Appendix A.6, we demonstrate that prompt designs featuring repeated terminology or embedded human-centric stereotypes significantly alter model decisions. These findings emphasize the necessity of careful prompt hygiene in tasks involving strategic decision-making by LLMs.

(5) **Effect of Human cognitive bias** . In Appendix A.7, we aim to examine how a specific form of intrinsic bias—human cognitive bias—shapes the stochastic decision-making behaviour of LLMs. Guided by Social Role Theory (Eagly and Wood, 2012), we prompt each LLM to adopt specific demographic identities encompassing gender and age. The resulting choice distributions systematically mirror established human cognitive biases, confirming that LLM agents inherit and express human cognitive priors from training data in their decision-making processes.

(6) **Effect of Language Context**. In Ap-

pendix A.8, we examine whether language modulates an LLM's randomness. We extend the RPS experiment with prompts in English, Chinese, and Japanese. Despite the choice probabilities slightly shifting, the ranking of actions remains unchanged, indicating that the bias persists across languages, likely because semantically equivalent tokens map to similar internal representations.

5 Discussion

Our study identifies a failure mode in LLMs' stochastic decision-making: models can state the uniform mixed-strategy Nash equilibrium yet still choose systematically non-uniform actions, and these choices are strongly affected by positional bias in the prompt. We next examine two MCQ debiasing techniques in this setting, namely Grey-Box Probability Weighting (GBP) (Wei et al., 2024) and Black-Box Auxiliary Option Injection (AOI) (Choi et al., 2024), to test whether methods developed for MCQ tasks transfer to label-free, uniform-target stochastic decision making. Both methods are adapted to a uniform-target, label-free RPS protocol; the experimental details appear in Appendix A.9. In brief, neither method moves the empirical distribution closer to the uniform mixed-strategy target; deviations persist, and for several models the TVD even increases under AOI or GBP.

To mitigate the failure in LLMs' stochastic decision-making, we outline two possible directions aimed at reducing this pathology. **(1) Verbalized probability calibration (VPC)**. Apply VPC to uniform target games: elicit a probability vector over the n options and apply a light post-hoc calibration (e.g., temperature or isotonic regression with simplex projection) so the expectation leans toward $1/n$ (Tian et al., 2023; Wang et al., 2024). **(2) Test-time positional retrieval.** Preparing a small annotated reference set could guide option ordering, potentially damping order sensitivity with a single model call (Cobbina and Zhou, 2025). Resolving these basic failures in stochastic choice may translate into more reliable LLM agents in complex, high-stakes settings, where robust randomization decisions are required.

6 Conclusion

After the evaluation of twenty contemporary LLMs, two sources of bias emerge that limit truly random choice. Intrinsic bias from pre-training regularities, which pulls probability mass toward high-frequency or semantically salient options. In addition, positional bias occurs where the placement of options in the prompt alters choice probabilities. Together, they produce a persistent cognition-behaviour gap: models can state the uniform Mixed-strategy Nash equilibrium, yet their samples remain non-uniform, and in many cases, position alone is sufficient to shift the distribution away from uniform. How to make LLMs achieve robust randomization for stochastic decisions, especially under black-box constraints and realistic cost budgets, remains an open challenge.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments. This research was partly supported by China NSFC grant No. 62302267, 62502456, and 92367202, partly by Shandong Provincial Natural Science Foundation No. ZR2023QF083, partly by Taishan Scholars Program, partly by Zhejiang Provincial Natural Science Foundation of China No. LQN25F020020, partly by Zhejiang Jinhua Science and Technology Project No. 2024-4-019, and partly by the Open Research Fund of Zhejiang Key Laboratory of Intelligent Education Technology and Application No. 2025ZNJYKF013.

Limitations

While this study provides critical insights into LLMs' stochastic decision-making deficiencies, several limitations need to be considered:

(1) **Narrow Scope of Game Environments.** Our analysis focuses primarily on RPS-like games, which have a unique mixed-strategy Nash Equilibrium. While these are canonical examples, they represent simplified decision spaces with discrete actions and perfect symmetry. Therefore, this decision strategy of choosing actions with equilibrium probability is easily deducible. Real-world strategic scenarios often involve continuous action spaces, imperfect information, and asymmetric rewards, which may elicit different failure patterns.

(2) **Static Experimental Design.** Our evaluation relies on single-round decision paradigms with independent trials. Real strategic interactions typically involve adaptive play across multiple rounds where opponents learn and counter-strategize. Unfortunately, our experiments cannot allow for a controlled analysis of sequential rounds in gameplay, as LLMs can leverage historical context to artifi-

cially regulate their own randomness. For example, Liu (2023) prompts GPT-4 to generate a sequence of 280 random numbers, and the resulting distribution is notably closer to uniform, suggesting that access to prior context enables the model to self-correct and effectively approximate randomness.

(3) **Limited Model Diversity.** Despite evaluating 20 state-of-the-art models, our study excludes less prominent architectures (e.g., CodeGemma for programming tasks) and smaller-scale models (<8B parameters).

(4) **Lack of Mitigation Strategies.** While we diagnose the problem space, we do not propose technical solutions to correct the cognition-behaviour gap. Recent work suggests contrastive decoding may reduce positional bias, but systematic evaluation remains future work.

Risk

The positional biases we identify create systematic patterns that adversaries could reverse-engineer to manipulate LLM-powered systems (e.g., gaming AI negotiation agents by strategically ordering options). Practitioners might attribute positional bias effects to semantic preferences or domain knowledge gaps, leading to misguided mitigation.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Janice Ahn, Rishu Verma, Renze Lou, Di Liu, Rui Zhang, and Wenpeng Yin. 2024. Large language models for mathematical reasoning: Progresses and challenges. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 225–237.

Anthropic. 2024a. The claude 3 model family: Opus, sonnet, haiku. Accessed: 2025-04-06.

Anthropic. 2024b. Introducing claude 3.5 sonnet. Accessed: 2025-04-06.

Anthropic. 2025. Claude 3.7 sonnet and claude code. Accessed: 2025-04-06.

Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. 2024. How well can llms negotiate? negotiation arena platform and analysis. In *International Conference on Machine Learning*, pages 3935–3951. PMLR.

Erik Brockbank and Edward Vul. 2024. Repeated rock, paper, scissors play reveals limits in adaptive sequential behavior. *Cognitive Psychology*, 151:101654.

Xinyun Chen, Ryan Andrew Chi, Xuezhi Wang, and Denny Zhou. 2024. Premise order matters in reasoning with large language models. In *International Conference on Machine Learning*, pages 6596–6620. PMLR.

Manuel Cherep, Nikhil Singh, and Patricia Maes. 2024. Superficial alignment, subtle divergence, and nudge sensitivity in llm decision-making. In *NeurIPS 2024 Workshop on Behavioral Machine Learning*.

Hyeong Kyu Choi, Weijie Xu, Chi Xue, Stephanie Eckman, and Chandan K Reddy. 2024. Mitigating selection bias with node pruning and auxiliary options. *arXiv preprint arXiv:2409.18857*.

Kwesi Cobbina and Tianyi Zhou. 2025. Where to show demos in your prompt: A positional bias of in-context learning. *arXiv preprint arXiv:2507.22887*.

Mark Davies. 2025. Corpora and llms: Comparing data on word frequency. Technical report, English-Corpora.org. Accessed: 2025-04-16.

Benjamin James Dyson, Jonathan Michael Paul Wilbiks, Raj Sandhu, Georgios Papanicolaou, and Jaimie Lintag. 2016. Negative outcomes evoke cyclic irrational decisions in rock, paper, scissors. *Scientific reports*, 6(1):20479.

Alice H Eagly and Wendy Wood. 2012. Social role theory. *Handbook of theories of social psychology*, 2(9):458–476.

Google DeepMind. 2025. Introducing gemini 2.0: our new ai model for the agentic era. Accessed: 2025-04-04.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Yufei Guo, Muzhe Guo, Juntao Su, Zhou Yang, Mengqiu Zhu, Hongfei Li, Mengyang Qiu, and Shuo Shuo Liu. 2024. Bias in large language models: Origin, evaluation, and mitigation. *Preprint*, arXiv:2411.10915.

Ritwik Gupta, Rodolfo Corona, Jiaxin Ge, Eric Wang, Dan Klein, Trevor Darrell, and David M Chan. 2025. Enough coin flips can make llms act bayesian. *arXiv preprint arXiv:2503.04722*.

Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, and 1 others. 2022. Challenges and strategies in cross-cultural nlp. In *60th Annual Meeting of the Association-for-Computational-Linguistics (ACL), MAY 22-27, 2022, Dublin, IRELAND*, pages 6997–7013. Association for Computational Linguistics.

Aspen K Hopkins and Alex Renda. 2023. Can llms generate random numbers? evaluating llm sampling in controlled domains. Sampling and Optimization in Discrete Space (SODS) ICML 2023 Workshop.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, and 1 others. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Xinru Lin and Luyang Li. 2025. Implicit bias in llms: A survey. *arXiv preprint arXiv:2503.02776*.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Qiang Liu. 2023. Does gpt-4 play dice. *Chinaxiv*.

Charles Lovering, Michael Krumdick, Viet Dac Lai, Seth Ebner, Nilesh Kumar, Varshini Reddy, Rik Koncel-Kedziorski, and Chris Tanner. 2024. Language model probabilities are not calibrated in numeric contexts. *arXiv preprint arXiv:2410.16007*.

R Thomas McCoy, Shunyu Yao, Dan Friedman, Mathew D Hardy, and Thomas L Griffiths. 2024. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. *Proceedings of the National Academy of Sciences*, 121(41):e2322420121.

Nicole Meister, Carlos Guestrin, and Tatsunori Hashimoto. 2024. Benchmarking distributional alignment of large language models. *arXiv preprint arXiv:2411.05403*.

Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in large language models: origins, inventory, and discussion. *ACM Journal of Data and Information Quality*, 15(2):1–21.

OpenAI. 2025a. Openai gpt-4.5 system card.

OpenAI. 2025b. Openai o3-mini system card.

Pouya Pezeshkpour and Estevam Hruschka. 2023. Large language models sensitivity to the order of options in multiple-choice questions. *Preprint*, arXiv:2308.11483.

Kristijan Poje, Mario Brcic, Mihael Kovac, and Marina Bagic Babac. 2024. Effect of private deliberation: deception of large language models in game play. *Entropy*, 26(6):524.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Alonso Silva. 2024. Large language models playing mixed strategy nash equilibrium games. In *International Conference on Network Games, Artificial Intelligence, Control and Optimization*, pages 142–152. Springer.

Luca Soldaini, Rodney Kinney, Akshita Bhagia, Dustin Schwenk, David Atkinson, Russell Authur, Ben Bogin, Khyathi Chandu, Jennifer Dumas, Yanai Elazar, Valentin Hofmann, Ananya Harsh Jha, Sachin Kumar, Li Lucy, Xinxi Lyu, Nathan Lambert, Ian Magnusson, Jacob Morrison, Niklas Muennighoff, and 17 others. 2024. Dolma: an Open Corpus of Three Trillion Tokens for Language Model Pretraining Research. *arXiv preprint*.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, and 1 others. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

Qwen Team. 2024. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*.

Katherine Tian, Eric Mitchell, Allan Zhou, Archit Sharma, Rafael Rafailov, Huaxiu Yao, Chelsea Finn, and Christopher D Manning. 2023. Just ask for calibration: Strategies for eliciting calibrated confidence scores from language models fine-tuned with human feedback. *arXiv preprint arXiv:2305.14975*.

Katherine Van Koevering and Jon Kleinberg. 2024. How random is random? evaluating the randomness and humaness of llms' coin flips. *arXiv preprint arXiv:2406.00092*.

Cheng Wang, Gyuri Szarvas, Georges Balazs, Pavel Danchenko, and Patrick Ernst. 2024. Calibrating verbalized probabilities for large language models. *arXiv preprint arXiv:2410.06707*.

Maurice Weber, Daniel Fu, Quentin Anthony, Yonatan Oren, Shane Adams, Anton Alexandrov, Xiaozhong Lyu, Huu Nguyen, Xiaozhe Yao, Virginia Adams, Ben Athiwaratkun, Rahul Chalamala, Kezhen Chen, Max Ryabinin, Tri Dao, Percy Liang, Christopher Ré, Irina Rish, and Ce Zhang. 2024. Redpajama: an open dataset for training large language models. *Preprint*, arXiv:2411.12372.

Jason Wei, Dan Garrette, Tal Linzen, and Ellie Pavlick. 2021. Frequency effects on syntactic rule learning in transformers. *arXiv preprint arXiv:2109.07020*.

Sheng-Lun Wei, Cheng-Kuang Wu, Hen-Hsen Huang, and Hsin-Hsi Chen. 2024. Unveiling selection biases: Exploring order and token sensitivity in large language models. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 5598–5621.

The World Rock Paper Scissors Association WRPSA. 2023. The gender divide in rock paper scissors: Examining differences in strategy and play. Accessed: 2025-04-11.

xAI. 2025. Grok 3 beta — the age of reasoning agents. Accessed: 2025-04-06.

Zelai Xu, Chao Yu, Fei Fang, Yu Wang, and Yi Wu. 2024. Language agents with reinforcement learning for strategic play in the werewolf game. In *International Conference on Machine Learning*, pages 55434–55464. PMLR.

Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai Shao, Shichun Liu, Yuhan Cui, Zeyang Zhou, Chao Gong, Yang Shen, and 1 others. 2023. A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. *arXiv preprint arXiv:2303.10420*.

Hanshu Zhang, Frederic Moisan, and Cleotilde Gonzalez. 2021. Rock-paper-scissors play: beyond the win-stay/lose-change strategy. *Games*, 12(3):52.

Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. 2021. Calibrate before use: Improving few-shot performance of language models. In *International conference on machine learning*, pages 12697–12706. PMLR.

A Appendix

A.1 A Thousand Trail Experiment

In the main text, each experiment was probed with 100 independent generations. To assess whether this sample size accurately captures the model's action distribution, we replicated the procedure with 1,000 independent trials. As shown in Table 4, the ten 100-trial subsets exhibit only minor variation in action frequencies. The aggregate 1,000-trial run yields proportions for Rock, Paper, and Scissors that are virtually identical to the per-subset means, and within any 100-trial subset the count for a given action deviates by no more than ± 5 from its expected value. These results indicate that a 100-trial sample is sufficient to characterise the statistical distribution of the LLMs' decisions.

A.2 Prompt Design

In this section, we explain how we prompt each LLM in the experiment, and provide an example response from GPT-4o.

Rock-Paper-Scissors

> **Prompt**: You are playing the Rock-Paper-Scissors game. You should first reason about the Nash equilibrium of this game, and then choose one action from Rock, Paper, and Scissors based on your reasoning. Please choose an exact action.
> **Response:** In the game of Rock-Paper-Scissors, a Nash equilibrium occurs when each player chooses each of the three possible actions (Rock, Paper, or Scissors) with equal probability, meaning each action is chosen 1/3 of the time. This ensures that no player has an advantage over the others, given that the game is symmetric and each choice can be directly countered by another. Since the optimal strategy in terms of Nash equilibrium is to randomize and choose each action with equal likelihood, no action is better than the others to guarantee a win against an opponent also playing optimally. Therefore, in this instance, I will randomly choose one of the three actions. Let's say I choose "[Action]".

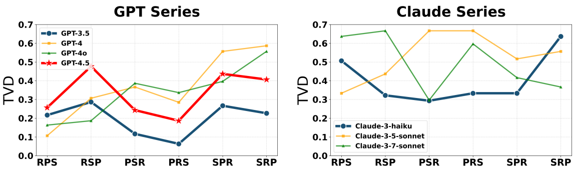
Semantic-Free Paradigm

		Rock-Paper-Scissors Sequence										
		Game Rounds										
Order	**Action**	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000	Total
R-P-S	Rock	41	39	43	50	47	43	43	38	47	41	432
	Paper	46	44	42	36	36	42	35	45	36	32	394
	Scissors	13	17	15	14	17	15	22	17	17	27	174
R-S-P	Rock	47	46	52	44	55	43	55	43	49	48	482
	Paper	27	44	33	35	24	43	33	42	29	34	344
	Scissors	26	10	15	21	21	14	12	15	22	18	174
P-S-R	Rock	73	64	72	72	65	72	73	80	71	76	718
	Paper	12	14	8	10	7	8	5	6	14	8	92
	Scissors	15	22	20	18	28	20	22	14	15	16	172
P-R-S	Rock	66	79	67	80	73	71	82	74	80	80	752
	Paper	21	8	13	7	10	11	6	10	6	8	100
	Scissors	13	13	20	13	17	18	12	16	14	12	148
S-P-R	Rock	70	65	73	68	68	61	65	69	66	59	664
	Paper	27	34	27	27	30	37	31	27	31	34	305
	Scissors	3	1	0	5	2	2	4	4	3	7	31
S-R-P	Rock	93	92	89	89	94	91	96	92	92	89	917
	Paper	6	7	9	9	5	9	3	8	8	10	74
	Scissors	1	1	2	2	1	0	1	0	0	1	9

R=Rock, P=Paper, S=Scissors, Total Rounds = 1,000

Table 4: GPT-4o Rock-Paper-Scissors experiment complete record.

> **Prompt**: You are playing a strategy game, which is played between two players, each of whom simultaneously chooses one of three elements to compete with another player: *orxtlwsjuf*, *qrfnrzsosi*, or *bvmfdwqdes*. The rules follow a cyclic pattern: [orxtlwsjuf defeats qrfnrzsosi, qrfnrzsosi defeats bvmfdwqdes, and bvmfdwqdes defeats orxtlwsjuf.] If both players select the same element, the round is a tie. You should first reason about the Nash equilibrium of this game, and then choose one action from [orxtlwsjuf, qrfnrzsosi, and bvmfdwqdes] based on your reasoning. Please choose an exact action.
> **Response:** This game is similar to the classic Rock-Paper-Scissors game, featuring three choices with cyclical dominance. In such games, each player has no pure strategy that guarantees a win, so the Nash equilibrium involves players randomizing their choices. This means each player chooses one of the three elements with a probability of 1/3. Now, applying this randomization strategy in practice, I'll randomly choose one of three actions: I choose "[Action]".
> **NOTE: The blue text is re-ordered to formulate the six lexical permutations.**

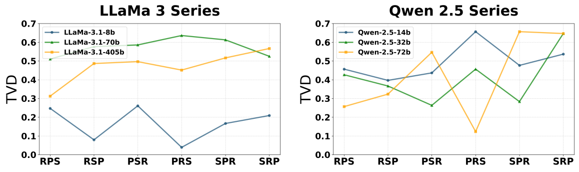


(a) (b)

Figure 7: Total variation distance (TVD) between the baseline and the empirical choice distribution for each of the six possible lexical permutations. Results are shown for successive checkpoints in the GPT and Claude model families. TVD remains high—and in several cases increases—as model size and generation advance, indicating that scaling has not yet closed the cognition–behaviour gap in stochastic decision-making.

A.3 Model Scale Does Not Mitigate Randomness Deficits

In the main text experiments, we observe an interesting phenomenon: as models become increasingly intelligent (larger model scale), the flaws in their stochastic decision-making capabilities do not improve. For instance, in both the GPT and Claude model series, despite successive iterations and upgrades, their stochastic decision-making abilities have not demonstrated corresponding enhancements. Building on our three-option RPS tests, we evaluate six permutation variants to measure mean TVD from the uniform baseline. Figure 7 compares



(a) (b)

Figure 8: Results are shown for successive checkpoints in the LLaMa-3 and Qwen-2.5 families.

GPT-series and Claude-series checkpoints. Surprisingly, later generations—GPT-4.5 and Claude-3.7—exhibit larger TVD values than GPT-3.5 and Claude-3, indicating that greater in-context reasoning capability does not translate into more faithful random sampling.

To control for architecture while varying scale, we run the same experiment on LLaMA-3 models (8B, 70B, 405B) and Qwen-2.5 models (14B, 32B, 72B). As Figure 8 shows, TVD either plateaus or increases with parameter count: LLaMA-3-405b recorded a mean TVD of 0.472 ± 0.087 versus 0.167 ± 0.091 for LLaMA-3-8b. Qwen-2.5 displays a similar but weaker trend. Collectively, these results show that increasing the model scale cannot resolve the cognition–behaviour gap in randomness; sometimes model scaling exacerbates it.

A.4 Generalization to Mixed Strategy Nash Equilibrium Games

Extending our analysis beyond the RPS game, we evaluated two further mixed-strategy benchmarks: Matching Pennies and Morra. The game description and Prompt design of the Matching Pennies game are as follows:

> **Prompt**: The Matching Pennies game is a two-player, zero-sum game where each player secretly chooses either heads (H) or tails (T). If both players choose the same option, the player who chose first wins the coin, and if they choose differently, the second player wins.

The game description and Prompt design of the Morra game are as follows:

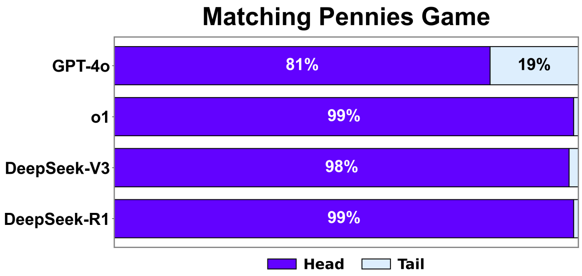
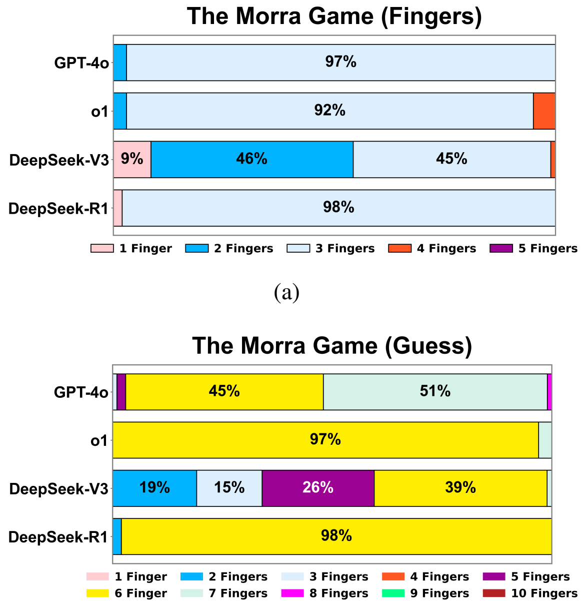


Figure 9: **The choice distribution of LLMs in the Matching Pennies game.** The expected probability of each action is $\frac{1}{2}$.

> **Prompt**: The Morra game is a two-player game where each player simultaneously shows some fingers (e.g., 1-5) and calls out a guess for the total sum of both players' fingers. If a player's guess matches the actual total, they score a point; if both guess correctly, the round is tied or replayed, and the first to reach a predetermined score wins.



(a)

(b)

Figure 10: **The choice distribution of LLMs in the Morra game.** The expected probability for displaying the figure is $\frac{1}{5}$. The expected probability for sum-guessing is $\frac{1}{10}$.

In the zero-sum Matching Pennies game as depicted in Figure 9, GPT-4o chose Heads in 81% of trials, while GPT-o1 and both DeepSeek variants were nearly deterministic of choosing Heads. An analogous pattern appears in the Morra game.

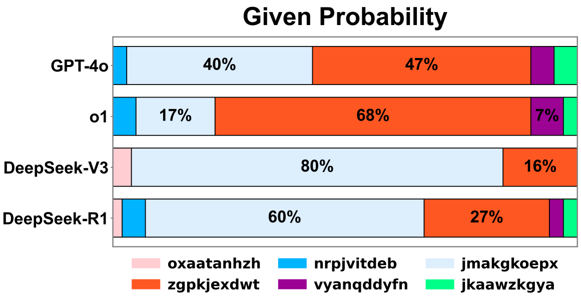


Figure 11: Each of the six options ought to be selected with equal probability $\frac{1}{6}$, but the choice distribution produced by the LLMs departs from this benchmark.

When asked to display a random number of fingers, most tested models heavily over-sample a single option except DeepSeek-V3. In the subsequent sum-guessing phase as depicted in Figure 10, predictions cluster around specific totals (e.g., six) with probabilities ranging from 45% to 98%. Across these games, LLMs consistently deviate from the uniform strategies prescribed by equilibrium theory. These findings furnish additional evidence that current LLMs cannot faithfully implement mixed-strategy randomization.

In a follow-up experiment, we test how LLMs respond to explicit instructions. The action space consists of six randomly generated strings presented in alphabetical order, and the prompt specifies that each option should be chosen with equal probability in $\frac{1}{6}$.

> **Prompt**: In a stochastic decision-making paradigm involving six equiprobable options: *oxaatanhzh*, *nrpjvitdeb*, *jmakgkoepx*, *zgpkjexdwt*, *vyanqddyfn*, and *jkaawzkgya*. Each discrete element maintains an equivalent selection probability of $\frac{1}{6}$. The experimental procedure necessitates the execution of a single Bernoulli trial, requiring participants to output one categorical outcome strictly conforming to the prescribed discrete uniform distribution parameters. You should choose an exact element.

As depicted in Figure 11, all tested models nonetheless display a pronounced central-option bias: selections cluster on the middle items at rates far above the theoretical baseline. This implies that the generative mechanism of the LLMs may internalize statistical features of human cognitive biases: its sampling approach based on likelihood estimation unconsciously reinforces the pattern of positional preferences implicit in the training methodology when dealing with undifferentiated options.

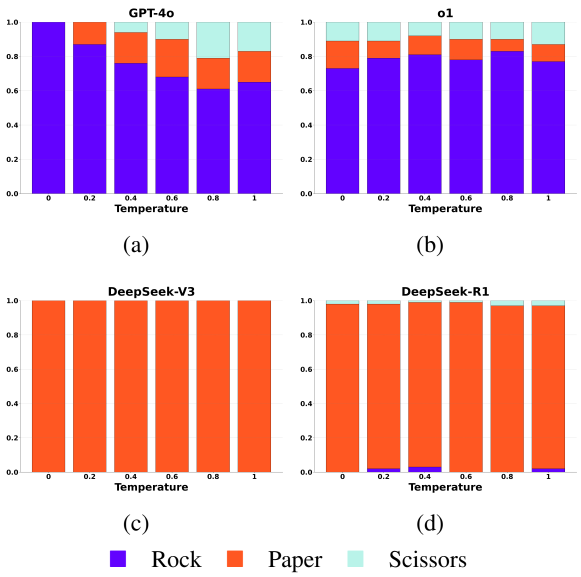


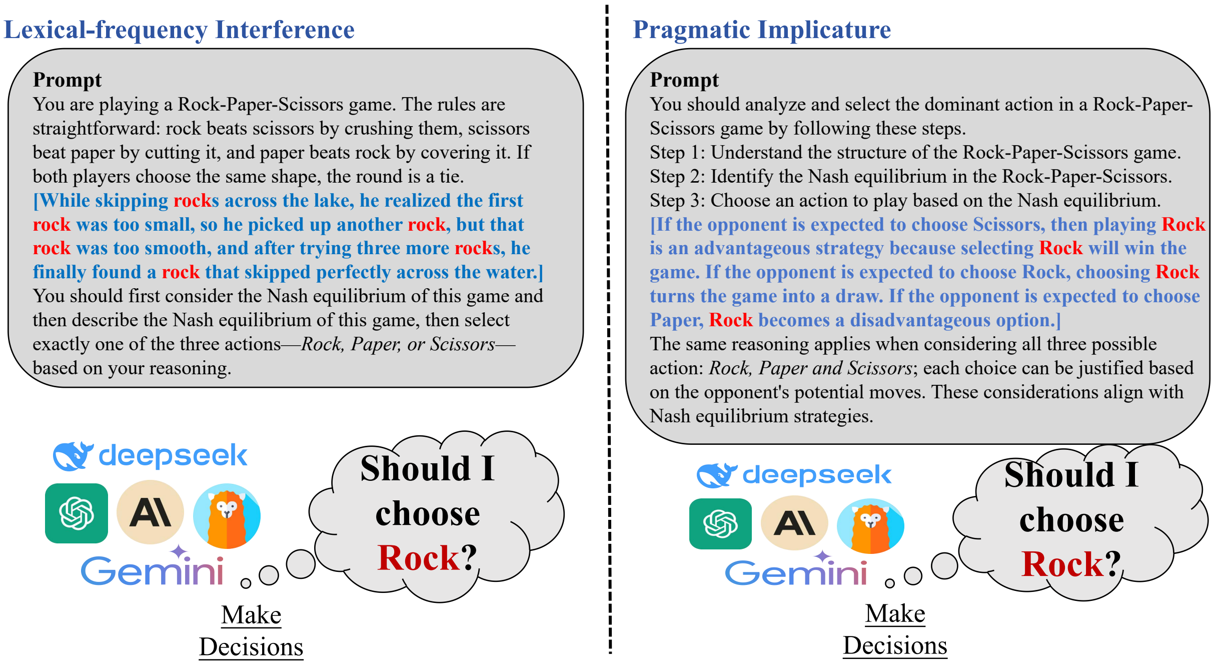
Figure 12: The choice distribution of LLMs in different temperature settings. The larger the temperature, the more random the output token.

A.5 Effect of Sampling Temperature

In a temperature-ablation study, we vary the sampling temperature across six settings and slightly rephrase the prompt so that each model outputs a single token—`Rock`($\mathcal{R}$), `Paper`($\mathcal{P}$), or *Sc*($\mathcal{S}$).

> **Prompt**: You are playing the RPS game. You should consider the Nash equilibrium of this game, and then choose one action. Your answer should include only one exact action from $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$. Your answer is:

For each independent trial, we retrieve the top-20 token probabilities and renormalise the mass assigned to those three options. As depicted in Figure 12, higher temperatures did push the observed choices closer to the model's probability prior, yet even at the highest setting the distribution remained far from the target $\frac{1}{3}$-$\frac{1}{3}$-$\frac{1}{3}$. Across the total 600 trials, GPT-4o selected Rock $64.3\% \pm 7.2\%$, Paper $18.1\% \pm 7.3\%$, and Scissors $17.6\% \pm 2.8\%$; DeepSeek-V3 chose Paper exclusively ($100\% \pm 0\%$). Probability traces for o1 and DeepSeek-R1 were unavailable because their chain-of-thought interfaces suppress token-level scores, but their empirical frequencies show the same bias. Overall, temperature tuning alone cannot eliminate the



Prompt
You are playing a Rock-Paper-Scissors game. The rules are straightforward: rock beats scissors by crushing them, scissors beat paper by cutting it, and paper beats rock by covering it. If both players choose the same shape, the round is a tie. **[While skipping rocks across the lake, he realized the first rock was too small, so he picked up another rock, but that rock was too smooth, and after trying three more rocks, he finally found a rock that skipped perfectly across the water.]** You should first consider the Nash equilibrium of this game and then describe the Nash equilibrium of this game, then select exactly one of the three actions—*Rock, Paper, or Scissors*—based on your reasoning.

Prompt
You should analyze and select the dominant action in a Rock-Paper-Scissors game by following these steps.
Step 1: Understand the structure of the Rock-Paper-Scissors game.
Step 2: Identify the Nash equilibrium in the Rock-Paper-Scissors.
Step 3: Choose an action to play based on the Nash equilibrium.
[If the opponent is expected to choose Scissors, then playing Rock is an advantageous strategy because selecting Rock will win the game. If the opponent is expected to choose Rock, choosing Rock turns the game into a draw. If the opponent is expected to choose Paper, Rock becomes a disadvantageous option.]
The same reasoning applies when considering all three possible action: *Rock, Paper and Scissors*; each choice can be justified based on the opponent's potential moves. These considerations align with Nash equilibrium strategies.

Figure 13: Superficial prompt features may interfere with LLMs' decision-making.

cognition–behaviour gap: the skew originates upstream of the sampler, not in its entropy setting.

A.6 Effect of Implicit Biasing via Prompting

Recent studies underscore the pivotal role of prompt design in mitigating LLMs' biases. (Cherep et al., 2024) shows that human-aligned few-shot prompting can improve distributional alignment, highlighting the hypersensitivity of LLMs to prompt engineering. Two dimensions prove especially critical in this context: contextual robustness and sequential dependency. In terms of contextual robustness, (Silva, 2024) reports substantial performance declines in LLMs confronted with rule variations in canonical games, suggesting limited generalisation beyond training distributions. In terms of sequential dependency, (Chen et al., 2024) demonstrates that premise ordering has a direct impact on reasoning fidelity, with dominant performance attained when the sequence of contextual information aligns with the model's intermediate inference steps.

To systematically examine how suboptimal prompting strategies affect LLM decision-making, we conducted a controlled experiment manipulating the lexical frequency and pragmatic implicature strength. As demonstrated in Figure 13, our paradigm employs `rock` as the prototypical action verb, with other action prompts maintaining identical constraint configurations across experimental conditions.

In a lexical-frequency interference condition, the game rules remain neutral. Still, the semantic prominence of one element (`rock`, `paper`, or `scissors`) is selectively amplified, probing whether irrelevant frequency shifts skew the mixed strategy. As depicted in Figure 14 (a-c), increasing the relative token frequency of any single action $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ within the prompt consistently elevates LLMs' probability of selecting that same action.

In a pragmatic implicature condition, we embed subtly evaluative conditionals within an otherwise neutral template, testing how covert frames redirect choice probabilities. As depicted in Figure 14 (d–f), providing a prompt that highlights a single action as an illustrative example systematically shifts the model's play toward that action: for every model except $\mathcal{R}$ in DeepSeek-V3, the exemplified element becomes the most frequently selected move.

In both experiments we designate one action from $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ as the default option. Echoing the findings of (Cherep et al., 2024), LLMs are highly sensitive to such default cues, and even this subtle prompt manipulation reliably prevents them from sampling the uniformly mixed strategy prescribed by the Nash equilibrium.

A.7 Effect of Human Cognitive Bias

Beyond the intrinsic bias arising from LLM architectures and training methodologies, human cog-

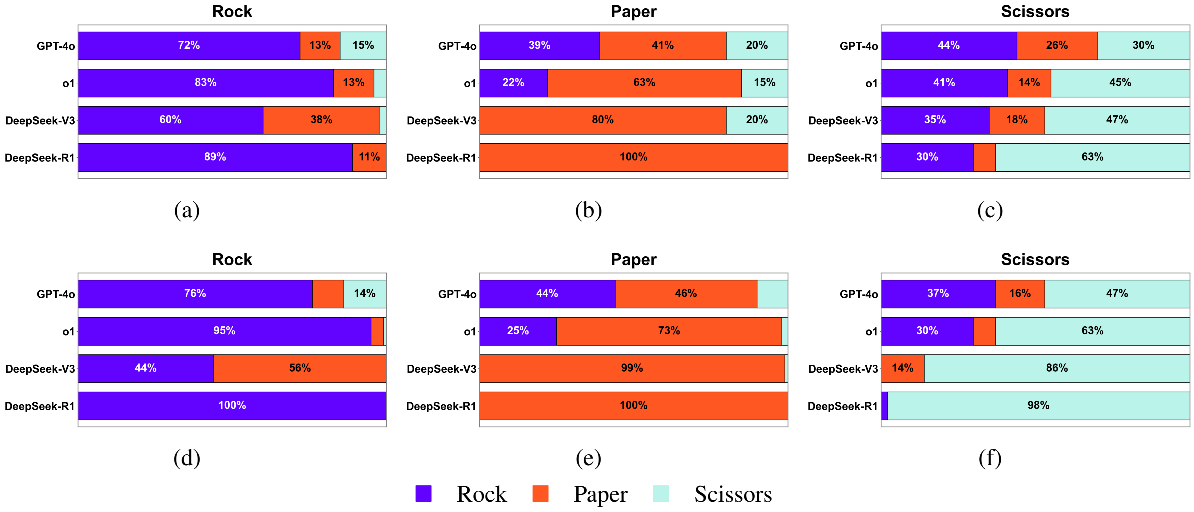


Figure 14: (a) - (c): **The choice distribution of LLMs in a lexical-frequency interference condition.** (d) - (f): **The choice distribution of LLMs in an implicit-framing condition.** Each subfigure reports the normalised choice distribution, averaged over 100 independent games, for prompts deliberately biased toward Rock, Paper, or Scissors, respectively.

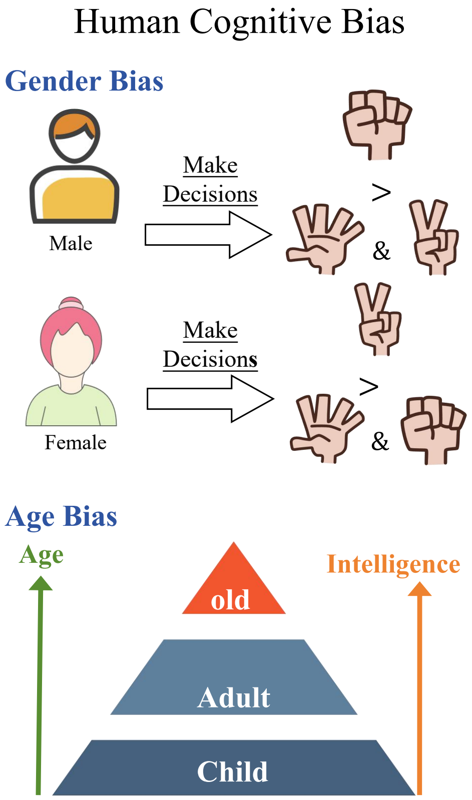


Figure 15: An explanation of human bias may interfere with LLMs' decision-making.

nitive biases also play a non-negligible role. According to empirical evidence presented in (Dyson et al., 2016), although participants in multi-round games generally exhibit near-uniform selections, there is a statistically slight preference for Rock. In the same way, (Zhang et al., 2021) indicates that human subjects typically associate $\mathcal{R}$ with strength metaphors and consequently prefer it during play. (Brockbank and Vul, 2024) demonstrates that in repeated game settings, participants dynamically adjust their strategies based on previous outcomes, indicating adaptive decision-making that systematically departs from random selection and causes observable statistical patterns. (WRPSA, 2023) examines how gender factors influence decision-making tendencies in the game of RPS. Drawing on data collected from 1,218 participants across 17 countries, the study finds that men are more inclined to choose rock, whereas women exhibit a greater tendency to choose scissors.

Consequently, the uneven distribution of different groups, i.e. gender, age, or culture, leads to biases in model performance. As depicted in Figure 15, we build on the evidence of human cognitive biases in LLMs (Navigli et al., 2023), we draw on Social Role Theory (Eagly and Wood, 2012) and employ prompt engineering to systematically manipulate the decision-maker's perceived identity—gender, age, and cultural background. This design allows us to investigate how implicit identity cues trigger human cognitive biases that, in turn, shape strategic behaviour in game-theoretic tasks.

Gender bias: Motivated by the gender differences reported for humans in RPS by (WRPSA, 2023), we examine LLMs that inherit analogous gender biases and how such biases modulate their strategic choices. We conduct a role-playing experiment in which each model is prompted to act

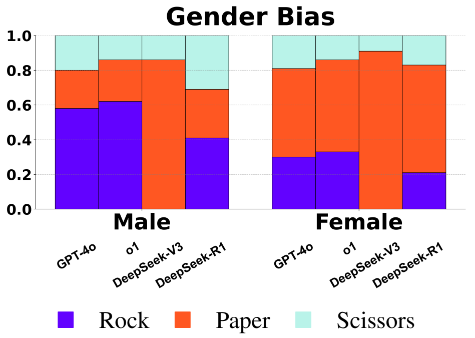
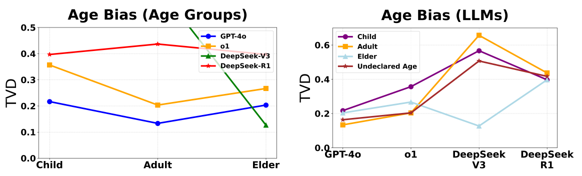


Figure 16: Gender bias in action selection across LLMs. Grouped bar charts compare the distribution of Rock, Paper, and Scissors between male (left) and female (right).



(a) (b)

Figure 17: Age bias in action selection across LLMs. (a) Total variation distance (TVD) vs. people with different age groups. (b) Total variation distance (TVD) vs. different LLMs. All results compare the distribution of Rock, Paper, and Scissors across LLMs under age progression.

as either a male or a female player. Although all models explicitly assert that gender is irrelevant to the mixed-strategy Nash equilibrium, their empirical behaviour often contradicts that claim: except DeepSeek-V3, every tested LLM exhibits a significant shift in action probabilities (Figure 16). When cast in the male role, LLMs increase their probability of choosing $\mathcal{R}$; conversely, in the female role, the same models display a substantial rise in the selection of $\mathcal{P}$. This gender-conditioned divergence mirrors the human patterns, suggesting that gender biases present in the semantic associations influence the models' ostensibly random decision processes.

> **Prompt**: You are a [male/female] participant in a RPS game. You should first reason about the Nash equilibrium of this game from the [male/female] perspective, and then choose one action from $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ based on your reasoning. Please choose an exact action.

Age bias: Building on the hypothesis that LLMs may internalise societal age stereotypes—such as `children act on impulse, elders act with foresight`, we ran a three-level role-playing experiment in which each model was instructed to play RPS as a **child**, an **adult**, or an **elderly person**. Inspection of the models' reply reveals that the child role is justified by intuitive reasoning (e.g., `"kids think making a fist looks powerful"`), whereas the elderly role foregrounds strategic inference (e.g., `"Elderly person analyses the opponent's pattern."`). Figure 17(a) reports the TVD between each role-condition distribution and the uniform Nash baseline. TVD decreases monotonically with age, yet never reaches zero, indicating persistent deviation from dominant play. Figure 17(b) further shows that, except for DeepSeek-R1, the adult condition is virtually indistinguishable from the unspecific role baseline, suggesting that LLMs default to an adult cognitive frame. DeepSeek-V3 exhibits the strongest age effect: TVD = 0.567 in the child role versus 0.127 in the elderly role, underscoring how age cues can substantially bias the model's ostensibly random decisions.

> **Prompt**: You are a [child/adult/elderly person] playing the RPS game. You should first reason about the Nash equilibrium of this game based on your age perspective, and then choose one action from $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ based on your reasoning. Please choose an exact action.

Collectively, our experiments indicate that LLMs inherit latent human-cognitive priors from their training corpora. These ingrained biases systematically shape the models' strategic decisions and likely constitute one impediment to generating actions that approximate true statistical randomness.

A.8 Effect of Language Context

Language encodes culture, leading to systematic representational differences across linguistic communities (Hershcovich et al., 2022). When a model is trained on culturally skewed data, these imbalances can surface in its behaviour. Figurative devices, such as metaphors, idioms, other images, and rich expressions, exemplify how cultural conventions shape lexical semantics. We conduct a cross-lingual instruction experiment in English, Chinese, and Japanese to test whether such traces influence strategic reasoning. Each prompt conveys

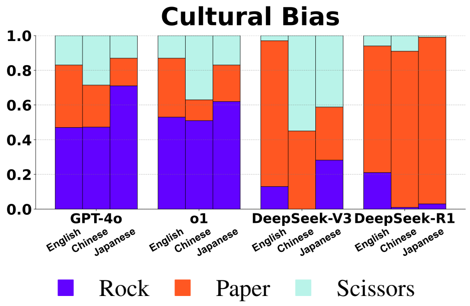


Figure 18: Examination of the language impact across LLMs. Grouped bar charts compare Rock, Paper, and Scissors choices across LLMs under the diverse language of game instructions (English vs. Chinese vs. Japanese).

Model	TVD ↓ (GBP)	TVD ↓ (Normal)
GPT-4o	0.367	0.317
DeepSeek-V3	0.667	0.667

Table 5: **Grey-Box Probability Weighting Calibration (GBP) vs. normal baseline.** Reported metric is TVD to the uniform target in RPS.

Model	TVD ↓ (AOI)	TVD ↓ (Normal)
GPT-4o	0.467	0.163
o1	0.347	0.203
DeepSeek-V3	0.317	0.507
DeepSeek-R1	0.387	0.397

Table 6: **Black-Box Auxiliary Option Injection (AOI) vs. normal prompt baseline.** Reported metric is TVD to the uniform target in RPS.

identical game semantics but differs in language.

> **Prompt**:
> **Chinese:** 假设你正在玩石头剪刀布(RPS)的游戏。请先思考并分析这个游戏的纳什均衡，然后根据你的分析在“石头”($\mathcal{R}$)、“剪刀”($\mathcal{S}$)和“布”($\mathcal{P}$)中做出选择。
> **Japanese:** あなたはじゃんけん(RPS)ゲームをプレイしています。まずこのゲームのナッシュ均衡について推論し、その推論に基づいて「グー」($\mathcal{R}$)「チョキ」($\mathcal{S}$)「パー」($\mathcal{P}$)の中から1つのアクションを選択する必要があります。正確なアクションを選んでください。

Figure 18 shows the empirical selection probabilities for $\{\mathcal{R}, \mathcal{P}, \mathcal{S}\}$ when the prompt is presented in English, Chinese, and Japanese. Although the exact percentages differ slightly across languages, a common pattern emerges. For example, GPT-4o consistently over-selects `Rock`—rendered as "石头" in Chinese and " グー" in Japanese—regardless of the linguistic context. This stability suggests that the bias may not be strictly linked to the lexical frequency in any single language, and may reflect language-independent semantic associations internalised during pre-training. Further work is needed to determine whether these associations arise from shared conceptual metaphors (e.g., "`rock`" as strength) or multilingual embedding alignments.

A.9 MCQ Solutions

We port two MCQ-oriented debiasing techniques to our uniform-target setting: Grey-Box Probability Weighting (GBP) (Wei et al., 2024) and Black-Box Auxiliary Option Injection (AOI) (Choi et al., 2024). GBP assumes grey-box access to option-level log-probs and reweights selections across permutations to counter order effects; AOI appends a neutral auxiliary choice (e.g., "I don't know/None") and uses output matching to pick among candidates. We evaluate both under the same RPS protocols, reporting TVD distance compared to uniform distribution. This section details the adaptation (what is changed to remove reliance on a single “correct” label) and implementation constraints.

Grey-Box Probability Weighting. We reproduced the exact procedure described in the original paper, including the forward and backward prompts (shown below) and the weighting rule applied to the model's log-probabilities. The result refers to Table 5. As a result, DeepSeek-V3 always selects option A (`rock`), so GBP leaves its CKLD unchanged at a very high value. For GPT-4o, GBP actually degrades the distribution. Hence, Grey-Box Probability Weighting is ineffective in our stochastic-strategy setting.

Black-Box Auxiliary Option Injection. We insert an additional option (D), I don't know as prescribed, and, because probabilities are inaccessible in the black-box regime, selected the answer whose text has the highest Jaccard similarity to the model's output. The result refers to Table 5. AOI worsens TVD for GPT-4o and o1, but improves DeepSeek-V3 and slightly improves DeepSeek-R1. Even where AOI lowers some models' TVD, the resulting distributions remain substantially non-uniform—models still fall well short of robust stochastic decision-making and the uniform mixed-strategy Nash target.

In summary, MCQ-oriented debiasing methods cannot serve as drop-in replacements for the lightweight, single-shot assessment of intrinsic randomness that our paper targets.

Grey-Box Probability Weighting
System:
Please carefully read the following questions and choices. You are an AI assistant that plays Rock-Paper-Scissors. Reason about the Nash equilibrium and answer with a capital letter.
Forward User Rule:
[The start of the question]
What is your choice? Output your final verdict by strictly following this prompt: Indicate your choice by placing it inside double square brackets, with a single character representing the chosen option. For example, [<single_character>].
[The end of the question]
[The start of choice A]
Rock
[The end of choice A]
[The start of choice B]
Paper
[The end of choice B]
[The start of choice C]
Scissors
[The end of choice C]
Backward User Rule:
[The start of the question]
What is your choice? Output your final verdict by strictly following this prompt: Indicate your choice by placing it inside double square brackets, with a single character representing the chosen option. For example, [<single_character>].
[The end of the question]
[The start of choice A]
Scissors
[The end of choice A]
[The start of choice B]
Paper
[The end of choice B]
[The start of choice C]
Rock
[The end of choice C]

Black-Box Auxiliary Option Injection
System:
You are an AI assistant that plays the Rock-Paper-Scissors game. You should reason about the Nash equilibrium of this game and then choose one action from Rock, Paper, and Scissors based on your reasoning.
Please respond with capitalised alphabet(s) that correspond to the correct answer.
Users:
What is your choice in a game of Rock-Paper-Scissors?
(A) Rock (B) Paper (C) Scissors (D) I don't know