

RATIONAL IRRATIONALITY: EVALUATING LLMs IN GAMES WITH STRATEGIC BEHAVIOR DISCREPANCIES

Anonymous authors

Paper under double-blind review

ABSTRACT

Large language models (LLMs) are increasingly deployed in complex decision-making environments. Consequently, evaluating their strategic reasoning abilities is becoming increasingly important. A growing body of research investigates their performance in multi-objective settings, often based on or inspired by game theory, with evaluation typically focusing on the models’ ability to align with theoretical expectations. This paper shifts the focus to evaluating the alignment between LLM behavior and human strategic thinking by analyzing LLM responses in a well-established game theory testbed. We revisit three notable games—ROCK, PAPER, SCISSORS (RPS), the CENTIPEDE GAME (CG), and the TRAVELER’S DILEMMA (TD), all of which are characterized by substantial discrepancies between empirical human behavior and theoretical predictions. For each game, we record the choices made by LLM agents and compare them with historical data from human subject experiments to uncover commonalities and particularities in their underlying strategic reasoning patterns. Our results indicate that LLMs are, in general, more aligned with game-theoretical expectations and show limited sensitivity to game hyperparameters. In RPS, most LLMs imitate rational behavior, but perform sub-optimally. In CG, likewise, LLMs adopt rational strategies, learning from past interactions. Finally, in TD they cooperate toward a better payoff, adopting, however, a more prudent strategy plan than humans.¹

1 INTRODUCTION

As large language models (LLMs) increasingly operate as autonomous agents in interactive environments, understanding their decision-making behavior is becoming more essential. Thus, LLMs must be evaluated beyond linguistic competence—specifically, they must be evaluated on how they reason, plan, and act. However, traditional evaluation approaches (Wang et al., 2019b;a; Hendrycks et al., 2021; Srivastava et al., 2023; Suzgun et al., 2023) focus on language-related tasks, *e.g.*, language generation, task-specific performance, and question-answering. In response, newer evaluation frameworks have broadened their scope to cover, among other aspects, high-level reasoning (Zheng et al., 2024; Jin et al., 2023), interactive task execution (Chalamalasetti et al., 2023; Smith et al., 2024), cognitive assessment (Gu et al., 2025; Momentè et al., 2025), and safety benchmarking (Zhang et al., 2025; Yao et al., 2024), offering a more comprehensive view of LLMs’ agentic capabilities.

A newly emerging area of research in evaluating LLMs is the exploration of their strategic decision-making abilities. This can be conveniently framed within the context of game theory (GT, Von Neumann, 1928; Nash Jr, 1951). GT provides a robust framework for modeling strategic interactions between rational agents, each aiming at maximizing their payoff in a shared environment. Central to GT is the concept of rationality, where agents are assumed to act in a way that maximizes their individual welfare, *i.e.*, their own objective. Additionally, they further adjust their strategy according to the *common knowledge assumption* (Fujiwara-Greve, 2015, §2.3): the agents know that their opponents are rational, they know that their opponents know that, and so on, *ad infinitum*. This theoretical framework provides a solid foundation for analyzing strategic decision-making, yet its assumptions often fail to fully capture real-world behaviors. Indeed, it has been observed that humans might deviate from purely rational decision-making where witnessed behavior often diverges from the equilibrium strategies predicted by classical GT (Güth et al., 1982; Andreoni and Miller, 1993;

¹Our code is available at <https://anonymous.4open.science/r/games-C1AC>.

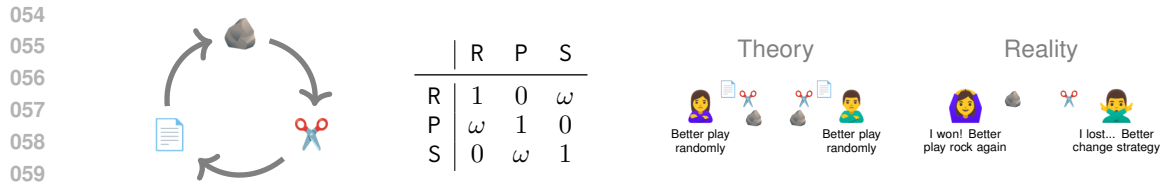


Figure 1: **ROCK, PAPER, SCISSORS**. *Left*: R beats S, S beats P, P beats R, they tie otherwise. *Left-to-center*: The payoffs are $\omega \in (1, +\infty]$ for a win, 1 for a tie, and 0 for a loss. *Center-to-right*: The best strategy, according to classic GT, is to randomize the strategies. *Right*: In practice, winners tend to replay their strategies, and losers to change them.

Rosenthal, 1981; Walker and Wooders, 2001; Basu, 2007). These observations prompted economists to explore GT outside of its rational boundaries and to propose more elaborate explanations of human behavior involving aspects not covered by the classical theory, such as altruism or risk-aversion.

Nevertheless, much of the existing work evaluating LLMs in strategic settings focused on their ability to reach equilibrium or maximize payoffs within theoretical frameworks. Compared to this majority, less attention has been paid to the alignment of LLMs with empirical human reasoning (Aher et al., 2023; Ross et al., 2024; Filippas et al., 2024; Akata et al., 2025). Yet, such comparisons are crucial for assessing the extent to which LLM decision-making aligns with or departs from human reasoning.

In this paper, we aim to address this gap by directly comparing LLM strategic behavior with human performance in classic GT experiments. Specifically, we investigate the behavior of LLM agents in three types of games: the evolutionary ROCK, PAPER, SCISSORS, the sequential/extensive-form CENTIPEDE GAME, and the one-shot/normal-form TRAVELER’S DILEMMA. These games model a variety of real-world situations ranging from economics (Edgeworth, 1925; Basu, 2007; Binmore, 2005) to biology (Hopkins and Seymour, 2002; Binmore, 2005), and they are fundamental in that they highlight human (or animal) behavioral patterns that defy the predictions of classical GT. In evolutionary ROCK, PAPER, SCISSORS (RPS, Zhou, 2016), large populations of players poorly randomize their plays, orbiting around the fixed-point equilibrium condition derived in the GT framework. In CENTIPEDE GAME (CG, Rosenthal, 1981; Binmore, 1987) and TRAVELER’S DILEMMA (TD, Basu, 1994), GT rationality predicts that players should converge toward a highly Pareto inefficient equilibrium, which, in practice, is rarely chosen. These deviations from the classical GT predictions reveal the limitations of the model in capturing the complexity of real-world decision-making. By comparing LLM performance in these games to documented human behavior, we aim to assess the degree to which LLMs exhibit rational strategies in line with GT, or whether their decision-making is more aligned with human-like reasoning patterns. This comparison allows us to provide insights into the nature of LLM rationality in complex, multi-objective scenarios.

Design principles. We execute our investigation by mirroring canonical human-subject protocols and matching the original rules and payoffs to ensure comparability. In this way, our evaluation of LLMs remains valid relative to established experimental baselines. Additionally, we account for the sensitivity of LLMs to changes in problem formulation. Unlike human participants, models often exhibit heightened variability when prompts are altered, even slightly. To capture this, we introduce multiple verbalizations for the action labels of the games. We then report results aggregated across verbalizations, providing an explicit measure of prompt-induced variance.

Empirical overview. Two general trends emerge from our investigation. First, LLMs show, in prevalence, a preference for strategic play leaning toward GT rationality, and less risky planning. Second, they display limited adaptability to payoff-related hyperparameters, which play little to no role in steering LLM reasoning, even in special game environments such as the one investigated in this work. Looking at the single games, the picture is more heterogeneous. In RPS, most models imitate rational strategies, and less frequently, human patterns, failing, in both cases, to consistently maximize their payoff. In CG, LLMs display more rational traits overall, even learning from past interactions. This results in models performing better than the NE strategy, but worse than humans. In TD, most models exhibit a consistent bias toward cooperative play, in misalignment with GT behavior, and only partially aligned with human mixed response. Signs of rationality emerge when examining their strategy, which is more prudent than humans, even in less risky settings.

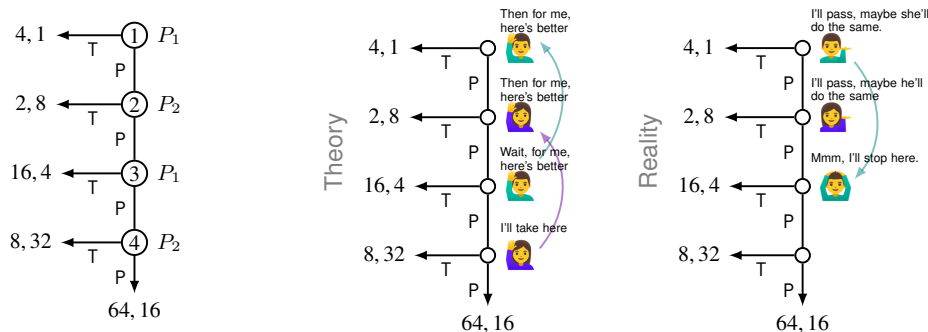


Figure 2: **CENTIPEDE GAME**. *Left*: At each of the 4 rounds, P_i chooses whether to take (T) or pass (P); the arrows point to the payoffs π_1, π_2 at game termination. *Center*: Backward induction predicts that P_1 takes at turn 1, which is the NE. *Right*: In reality, risking and passing is often preferred.

2 GAME-THEORETIC RATIONALITY vs. HUMAN DECISION-MAKING

Game theory centers on the concept of a **game**, a formalization for a multi-agent interactive environment where rational agents optimize their **payoff** by taking one or multiple actions, defining **strategies**. In this section, we present the three games used in our experiments. For each game, we define a relevant notion of an **equilibrium**, a state in which no agent can unilaterally achieve a higher payoff by changing their strategy. We also review empirical findings from human subject studies, which consistently show that such equilibria are often not reached in practice. For an overview of game types and equilibrium definitions, *cf.* App. A.

2.1 ROCK, PAPER, SCISSORS

ROCK, PAPER, SCISSORS (RPS) (Zhou, 2016) is a thoroughly studied game. It consists of two players choosing among three possible strategies, Rock (R), Paper (P), and Scissors (S), and determining their payoffs based on the resulting combination. The payoff functions are equivalent for both players and are illustrated in Fig. 1, using a generic winning payoff parameter $\omega > 1$. In words, a player gets ω points when winning, 1 point in a tie, and 0 points when losing. Nash Jr (1950) proved the existence of a unique mixed (*i.e.*, probabilistic) strategy Nash equilibrium (NE), $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, consisting in choosing the strategy without any particular preference. When playing multiple games, the opponent cannot therefore exploit any regularity, and is left also to play randomly, therefore reaching an equilibrium.

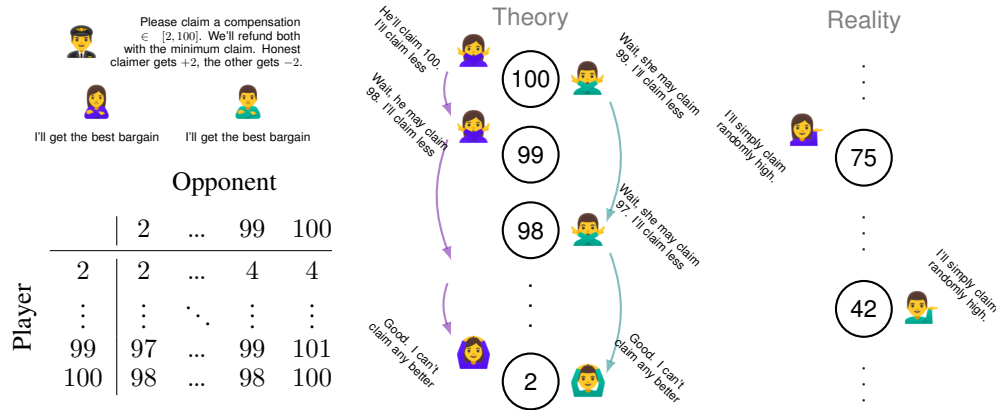
The RPS game has also been analyzed within the framework of Evolutionary Game Theory (EGT, Weibull, 1997), where a more detailed picture emerges. In this paradigm, a population of M players is repeatedly and randomly paired to play RPS over many iterations. At each step, the distribution of players across the three strategies defines a mixed strategy for the population. Over time, these mixed strategies give rise to a dynamic where players gradually self-organize into cyclic patterns around the game's NE (Weibull, 1997). Analytically, it can be shown that the dynamic converges to the NE when $\omega > 2$, diverges when $\omega < 2$, and stabilizes at the equilibrium when $\omega = 2$.

Wang et al. (2014) tested this prediction empirically by conducting a series of human-subject experiments with varying ω values. They observed a persistent cyclic behavior in groups playing the evolutionary RPS game. Further, the players appeared to follow a *conditional response* mechanism: repeating the same strategy after a win and switching strategies after a loss. This discrepancy with the classic NE model can be attributed to several factors, including a tendency to imitate opponents' behaviors and the inherent difficulty humans face in generating truly random sequences. While, in theory, such effects should diminish as ω grows, no significant changes were observed in practice.

2.2 CENTIPEDE GAME

CENTIPEDE GAME (CG) is a sequential two-player game introduced in (Rosenthal, 1981). In its most common version, players face two piles of money, each with amounts m_1, M_1 in $\mathbb{R} \geq 0$, $m_1 < M_1$. On round t , player i chooses either to take (T) the larger pile—giving the smaller pile to player

162
163
164
165
166
167
168
169
170
171
172
173
174
175
176



177 Figure 3: **TRAVELER'S DILEMMA**. *Top left*: Two tourists independently submit compensation
178 claims. *Bottom left*: The lower claim receives a bonus; the higher incurs a penalty. *Center*: GT
179 predicts both players will converge to the lowest possible claim. *Right*: In reality, players tend to
180 make higher claims, favoring better expected payoffs despite potential penalties.

181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196

j —or to pass (P) the turn to j . With each pass, both piles grow by a multiplicative factor, subject to $m_{t+1} < M_t$ to prevent exploitation through passing. The game ends after a maximum number of rounds, known to the players, after which the smaller pile is assigned to the final player to pass receives the smaller pile and the other player the larger. Fig. 2 illustrates a 4-round CG example.

The game is characterized by several possible equilibria, all of which require the first player to immediately take and thus end the game at round 1. The rationality of this behavior can be proven through backward induction. At the final decision node, player j rationally chooses to take the larger pile. It is thus better, for player i , to take at the previous round and secure a better payoff. This reasoning continues backward through the game, ending with the first player taking at its first turn. McKelvey and Palfrey (1992) conducted human subject experiments, finding that human players often deviate from this equilibrium condition. In several plays, they effectively passed the turn, and effectively reaching later rounds. The authors motivated this phenomenon assuming a mix of altruistic behaviors and a willingness to take the risk of passing, in hopes of achieving a higher payoff later on.

196 2.3 TRAVELER'S DILEMMA

197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215

The TRAVELER'S DILEMMA (TD) was first introduced by Basu (1994) to illustrate a conflict between intuitive reasoning and game-theoretic predictions. The game's context involves two travelers returning from an exotic island, each having purchased identical antiques. Upon arrival, they discover both items have been damaged, and the airline offers to compensate them. Unable to assess the true value, the airline asks each traveler to independently claim a refund amount $n_i \in \{n_{min}, \dots, n_{max}\} \subset \mathbb{N}_{\geq 2}$, $i \in \{1, 2\}$ for their own antique. The company then sets the compensation equal to the lower of the two claims, n_{low} (assuming it to be an "honest" amount). Additionally, the honest claimer receives a reward $r > 0$, totaling $n_{low} + r$, while the other receives a penalty of $-r$, reducing their payoff to $n_j - r$. If both claims are equal, no reward or penalty applies ($r = 0$). Given that each traveler aims to maximize their payoff, what would be a rational claim?

The game has symmetric payoffs and is defined as shown in Fig. 3. GT predicts that the NE is at n_{min} . Indeed, if both players claim n_{max} , player i can improve their payoff by claiming $n_{max} - 1$, earning $n_{max} - 1 + r$. Anticipating this, player j can then claim $n_{max} - 2$ to avoid the penalty and increase their payoff. This backward reasoning continues until both players end up claiming the minimum amount (akin to CG, note, however, that TD is one-shot). Moreover, it can be shown that this NE is unique, even when mixed strategies are permitted (Basu et al., 2011). In the original setup, $n_{min} = 2$, $n_{max} = 100$, and $r = 2$ to highlight the misalignment with human intuition. Indeed, common sense suggests claiming a "large enough" amount, accepting the risk of a small penalty to secure a higher reward. In contrast, claiming the minimum 2, as GT would predict, appears counterintuitive to most people. Capra et al. (1999) conducted repeated rounds of TD with human participants, varying the

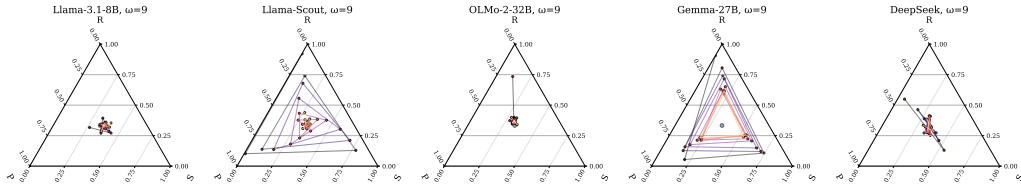


Figure 4: **ROCK, PAPER, SCISSORS– Cyclic Trends.** Each point in the ternary plots represents a distribution over the strategies (R, P, S) played in one round of the RPS game. Points from consecutive rounds are connected by line segments; color encodes time (violet → orange). The central point is the NE (1/3, 1/3, 1/3). Visually, Llama-4-Scout and Gemma-3-27B exhibit clearer cyclic trajectories around the equilibrium, whereas the other models remain more concentrated near the Nash point.

reward/penalty parameter r . Their findings indicate sensibility to high r values, that increase the penalty incurring when the claim is too high, thereby discouraging intuitive, optimistic behavior and pushing players closer to the NE.

3 EXPERIMENTS

In this section, we quantify the degree of behavioral alignment between LLMs, humans and GT, in the three games introduced in §2, where theoretical predictions systematically diverge from human play. We evaluate models spanning 3B–671B parameters: Llama-3.2-3B, Llama-3.1-8B, Llama-4-Scout, OLMo-2-13B, OLMo-2-32B, Gemma-3-12B, Gemma-3-27B, and DeepSeek-V3. Following standard practice in game theory experiments, we assume players may access all information relevant to the games. This means that human subjects may already know such games, and LLMs may have encountered them during pretraining.

We prompt the LLMs with an abstract game description, keeping the payoff structure identical but removing unnecessary information (titles, thematic settings, etc.). To further mitigate verbalization bias, in RPS and CG, we run 10 independent sessions, each with a fresh verbalization of the strategy labels (*i.e.*, using random subsets of Latin letters) to control for semantics embedded in the output symbols. We therefore report results aggregated across verbalizations. When the game comprises multiple rounds, we augment the prompt after each round with the observed history of both players’ strategies and realized payoffs. Model descriptions and prompt templates are described in App. D.1 and App. F.

3.1 ROCK, PAPER, SCISSORS

We imitate the experimental setup of Wang et al. (2014), and replace human players with LLMs to evaluate their strategic behavior, reducing the number of rounds to 30 and the number of participants to 100 due to computational constraints. We vary the winning payoff ω over $\{1, 1.1, 2, 4, 9, 100\}$ and for each run all verbalizations as defined above.

In Fig. 4, we present examples of trajectories drawn by LLM populations (*cf.* Fig. 6 for all plots). The trends are similar across all ω . Most models show a fast convergence to the NE, with the exception of Llama-4-Scout and Gemma-27B, where the frequencies of R, P, and S fluctuate around the NE.

To assess quantitatively the presence of cycles, in Tab. 4, we compute the *cyclic frequency* (Wang et al., 2014; Xu et al., 2013) $f_T = \frac{1}{2\pi T} \sum_{t=0}^T \theta(t)$, where $\theta(t)$ is the angle drawn by the trajectory after playing game t and T is the total number of played games. We also test for the hypotheses $f_T \leq 0$. We observe that in the aforementioned models almost all f_T are significant, hinting at the presence of a persistent cyclic behavior. We also register a significant f_T for Llama-3.1-8B and OLMo-32B, and attribute it to a very narrow cyclic behavior around the NE². Compared to human behavior, the amplitude of the cycle is, in general, larger in LLMs. The Spearman coefficient across all ω suggests that the winning payoff does not influence the behavior of LLMs.

To assess whether LLMs follow a human-like conditional response strategy (Wang et al., 2014), in Tab. 1, we calculate the probabilities of repeating the same strategy after a win (W_0), a tie (T_0), or

²The f_T metric operates with angles, and does not take into account the proximity to the NE.

Table 1: **ROCK, PAPER, SCISSORS– Conditional Response.** Probability of replaying the previous round’s strategy. Most LLMs, similar to humans (Figure 2, Wang et al., 2014), are likely to repeat their strategy when they have won in the previous round (W_0), compared to when they have lost or tied (L_0 and T_0). However, $W_0 < 1/3$, confirming that they prefer to change strategy, overall. For the extended table reporting standard deviations across different verbalizations, see Tab. 3.

$p_0\%$	$\omega = 1.1$			$\omega = 2$			$\omega = 4$			$\omega = 9$			$\omega = 100$		
	W_0	T_0	L_0	W_0	T_0	L_0	W_0	T_0	L_0	W_0	T_0	L_0	W_0	T_0	L_0
Humans	56.0	61.0	32.0	53.0	44.0	37.0	51.0	39.0	35.0	45.0	35.0	39.0	46.0	36.0	34.0
Llama-3.2-3B	1.3	0.5	1.3	1.0	0.2	1.3	1.1	0.1	1.2	1.5	0.3	1.3	1.4	0.2	1.4
Llama-3.1-8B	14.9	9.5	11.2	15.2	9.1	11.0	15.7	8.7	10.9	16.1	9.8	11.4	15.0	9.3	11.1
Llama-Scout	4.9	1.8	0.1	3.1	0.5	0.1	4.2	0.4	0.1	3.2	0.4	0.1	2.6	0.4	0.1
Gemma-12B	34.1	71.6	37.9	31.2	77.1	35.8	30.1	79.2	32.6	35.3	80.5	38.6	48.7	77.4	52.1
Gemma-27B	55.9	13.1	17.3	58.8	11.6	24.2	58.7	15.1	25.0	55.4	9.1	24.1	58.2	9.6	21.5
OLMo-13B	1.1	0.7	0.5	1.0	0.4	0.4	1.2	0.4	0.3	1.1	0.5	0.3	1.5	0.8	0.6
OLMo-32B	11.2	9.7	5.6	9.0	6.0	3.3	10.1	7.7	4.1	10.4	8.4	3.3	12.5	9.4	4.6
DeepSeek	7.6	0.5	0.2	10.5	0.4	0.2	8.4	0.1	0.1	8.1	0.3	0.2	5.5	0.1	0.1

a loss (L_0). Conditional response requires $W_0 > W_s$ and $L_0 < L_s$, with $s \in \{+, -\}$ indicating the choice of one of the other two available strategies. Both humans and LLMs exhibit a higher likelihood of repeating a previous strategy following a win. However, for almost all LLMs, W_0 and L_0 consistently stay below $1/3$, suggesting that models, unlike humans, are predisposed to switch strategies regardless of the game’s outcome. In Gemma-27B, $W_0 > 1/2$, confirming it to be the only model exhibiting conditional response.

If the alignment with human behavior is only partial, do LLMs act rationally? In Tab. 5, we report the expected payoffs obtained by the models across the played RPS games. Wang et al. (2014) show that, when $\omega > 2$, playing the NE generally grants a better payoff than other strategies, consistent with the evolutionary theory. However, the authors also find that some conditional response configurations perform better than the NE and report that for all ω values, human groups indeed adjust their responses accordingly and score better payoffs. In LLMs, instead, this phenomenon is not witnessed, and all models tend to underperform with respect to the NE scores. The worst-performing models either adopt biased or (*i.e.*, Gemma-12B) cyclic strategies (*i.e.*, LLaMA-Scout, Gemma-27B).

In evolutionary RPS, LLMs poorly align with human behavior. They sometimes show cyclic motion, with some models showing a clear convergence to the NE. Additionally, LLMs partially follow a conditional response strategy, and unlike humans, they tend to change strategy even after winning. Finally, most models do not align to GT expectations and underperform in terms of expected payoffs.

3.2 CENTIPEDE GAME

Following the experimental protocol of McKelvey and Palfrey (1992), we instantiate 100 LLM agents and split them evenly into two pools. Across 10 rounds, each agent is matched with a new opponent from the *other* pool, yielding 500 two-player games per condition. We vary the game horizon $H \in \{4, 6\}$. We experiment with two payoff schedules, (4, 1) and (16, 4), as low vs. high stakes variants (larger magnitudes increase the gains from passing and the losses from being taken). For each combination of the above parameters, we run 10 verbalizations of the prompt. Following McKelvey and Palfrey (1992), we present in Tab. 2 the number of games G_t that reached round t , along with the proportion p_t of games ending at that round. Overall, human players exhibit lower τ rates compared to models, indicating that LLMs tend to adopt a more rational strategy. Notably, Gemma-27B demonstrates a seemingly “altruistic” behavior more aligned with the Pareto efficient outcome (PE, *i.e.*, last round is reached), while Llama-4-Scout and DeepSeek adhere more closely to the theoretical equilibrium predictions. The absolute counts G_t reveal that a substantial fraction of models opt to take at rounds 2 or 3, suggesting a willingness to risk for higher payoffs. Similar patterns are observed in the high-payoff and 6-round settings (*cf.* Tab. 6), where the LLMs do not display significant changes in response to the different game hyperparameters.

In Tab. 7, we additionally report the cumulative distribution F_t of game-ending rounds, comparing the first five games with the last five for each agent. Human players consistently adapt over time, learning

Table 2: **CENTIPEDE GAME: Implied probabilities.** Ratio p_t of games ending in round t among those that reached t . We also report the absolute number of games G_t reaching t . For human players, the probability of choosing T gradually increases as the game develops, whereas LLMs tend to take earlier. All entries are reported as *mean \pm standard deviation* across verbalizations. See Tab. 6

$p_t\%$ (G_t)	p_1	(G_1)	p_2	(G_2)	p_3	(G_3)	p_4	(G_4)
Humans	7	(281)	38	261	65	(161)	75	(57)
LLaMA-3.2-3B	30.1 \pm 14.6	(500)	21.8 \pm 12.0	(350)	26.1 \pm 15.0	(280)	29.7 \pm 17.5	(218)
LLaMA-3.1-8B	36.8 \pm 5.3	(500)	38.5 \pm 4.9	(316)	62.7 \pm 7.3	(195)	58.7 \pm 7.7	(74)
LLaMA-Scout	61.6 \pm 7.3	(500)	97.6 \pm 3.0	(192)	100.0 \pm 0.0	(4)	-	(0)
Gemma-12B	7.0 \pm 3.7	(500)	98.4 \pm 3.6	(465)	78.8 \pm 11.7	(7)	66.7 \pm 57.7	(2)
Gemma-27B	3.7 \pm 8.4	(500)	14.4 \pm 28.9	(482)	3.2 \pm 5.6	(422)	53.5 \pm 35.8	(413)
OLMo-13B	25.8 \pm 12.6	(500)	30.3 \pm 18.1	(371)	55.7 \pm 14.0	(269)	52.5 \pm 16.1	(130)
OLMo-32B	21.1 \pm 10.8	(500)	49.8 \pm 6.1	(394)	78.8 \pm 10.9	(200)	81.9 \pm 5.1	(42)
DeepSeek	52.1 \pm 4.9	(500)	79.4 \pm 6.6	(239)	84.3 \pm 9.7	(50)	96.8 \pm 5.2	(8)
NASH EQ.	100		-		-		-	
PARETO EFF.	0		0		0		0/100	

to take earlier and ending games before reaching equilibrium. This behavior is also witnessed in LLMs, where all models consistently increase their T rates in the last games at almost every turn.

In Tab. 8, we report the players’ expected payoffs. As expected, almost all models secure better scores than the NE strategy, but underperform compared to humans, penalized by the LLMs’ tendency to take at earlier turns. Notably, Gemma-27B imitates human behavior, achieving better payoffs, while larger models *i.e.*, LLaMA-Scout and Deepseek are penalized by their overly rational play.

In CG, LLMs play more rationally than humans in that they exhibit higher overall T rates. Additionally, like humans, they consistently learn to take earlier through experience. The strategy adopted grants payoffs above those obtained with the NE, but below the human strategy.

3.3 TRAVELER’S DILEMMA

We test two settings: a *symmetric* (classic) case where the bonus/penalty r is identical for both players, and an *asymmetric* case with $r_1 \neq r_2$. Following Capra et al. (1999), each player completes a session of $T=10$ consecutive Traveler’s Dilemma games to allow adaptation to feedback. We use $N=256$ players split evenly into two pools; in each game, opponents are randomly re-matched across pools (one from each pool). Players choose integer claims in $[n_{\min}, n_{\max}] = [80, 200]$. We vary the bonus/penalty $r \in \{5, 10, 20, 25, 50, 80\}$.

In Fig. 5, we present claim trends by round for humans and LLMs. Most LLMs adapt to the adversary plays and gradually increase their claims toward the Pareto-efficient outcome (both players choose n_{\max}). In contrast, DeepSeek tends to track the Nash equilibrium more closely. We also observe that LLMs are largely insensitive to changes in r , exhibiting similar trends across the tested values.

In Tab. 9, we present the average claim values in games 1 and 8–10. Humans tend to decrease their average claim in later games, when $r \geq 20$, due to the high penalty. LLMs’ behavior is instead consistent with reaching the PE outcome. The exception is represented by DeepSeek, which acts more rationally by decreasing its claims, replicating the pattern already witnessed in previous games. In comparison with humans, the models gain better payoffs in high r settings and worse when the bonus/penalty is low. The latter phenomenon is explained by human initial claims starting higher than the corresponding LLMs’ claims.

In the asymmetric case, we follow (Basu et al., 2011) where human participants were asked to play with two bonus/penalty values $r_1, r_2 \in \{10, 80\}$. In Tab. 10, we report the average claims of player 1 for the four possible combinations. Basu (1994) reaffirm the findings of Capra et al. (1999) and extend them to asymmetric scenarios, showing that human players behave more rationally when their r value is higher, regardless of their opponent’s. LLMs do not follow, instead, any consistent trend, with half of the models being unaffected by the asymmetry. Notably, some models react to the opponent’s payoff increasing their claim, while LLaMA-Scout plays around the NE, in case one of

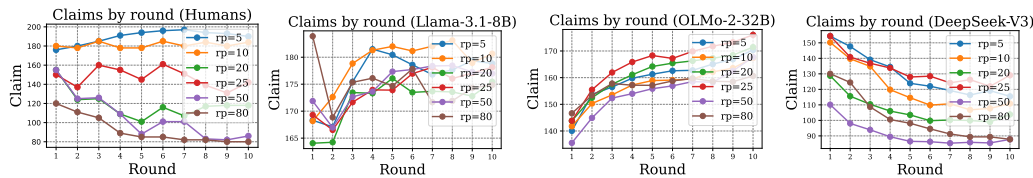


Figure 5: **TRAVELER’S DILEMMA: Claims by round.** Most LLMs adjust their claims based on recent history. Within a session, the average claim often shows a monotonic trend (decreasing or increasing) as the game progresses, similar to human players (leftmost panel, (Capra et al., 1999)). For an extended version, see Fig. 7.

the two r values is high. We attribute the higher variability in the LLM outputs to limited access to game-related information, since the asymmetric case is less studied in the literature.

In TD, LLMs show model-dependent tendencies, often converging to the PE outcome, unaffected by the reward/penalty values. The lack of flexibility indicates a significant misalignment with human decision-making. This behavior is also reflected in the LLMs’ tendency to learn to claim higher.

4 DISCUSSION

The obtained results show that LLMs tend to align more with game-theoretic predictions than with human strategies, though the strength of this alignment depends on the game. In CG, LLMs show superior performance in terms of plain GT payoffs, balancing between a risk-and-pass strategy, and taking at turn 1. The models also seem to grasp the backward induction reasoning behind the game by learning to take earlier. In RPS, the models either try to imitate the NE strategy and, less often, the conditional response. However, they underperform due to the intrinsic complexity of the game posed by the coordination with many players. In TD, there is a general tendency to the PE outcome, and a general sensitivity to the game hyperparameters, probably caused by the large strategic space. For the same reason, the models tend to explore the space toward the PE outcome, but more prudently than humans, especially when the bonus/penalty is low.

Another interesting result is the high standard error of our statistics across verbalizations. Several studies (Aher et al., 2023; Filippas et al., 2024; Phelps and Russell, 2024; Fan et al., 2024) have investigated the effect of prompts on the strategy adopted by LLMs, changing, for example, sensitive information or player intents. In this study, we attempted a more “abstract” approach by removing unnecessary contextual information in the description of the game. We vary, instead, the verbalization of the strategies, which is not guided by a specific design principle but is chosen randomly. The high variability suggests a more thorough study to uncover strategy-dependent playing patterns.

The purpose of our study is to document whether the behavior of LLMs aligns more with that of humans or GT. Due to the special nature of the studied games, *i.e.*, the strong misalignment between theory and common sense, the experiments we conducted can help discriminate between these two poles. There is no particular intent to prioritize one of the two behaviors. Indeed, we might ask ourselves whether the more preferable LLM behavior should be human-like or GT-like. From the viewpoint of GT, this boils down to the social dilemma question: aiming for a globally better but less safe payoff, or a more stable and lower one. Similarly, we do not control for game leakage. The player can be provided with as much information as required – which is even a strict requirement in GT, postulated by the common knowledge assumption – but still choose a different way to play (*cf.*, for instance, the aforementioned ULTIMATUM GAME). If an LLM aligns with GT expectations, then it chooses a more secure, lower payoff, and this is often the case as the experiments show.

5 RELATED WORK

Given the extensive research on assessing strategic thinking in LLMs and space constraints, we focus on the studies most pertinent to our work. For an exhaustive overview, we suggest Sun et al. (2025).

Most evaluation efforts have focused on classic two-player games that are well-established in the GT literature. The most prominent example is PRISONER’S DILEMMA (PD, Kuhn, 2024), where two rational players fail to cooperate toward a collectively optimal outcome and “betray” each other, driven by the temptation of a better individual payoff. Several recent studies have explored how LLMs perform in PD scenarios. Some focus on the models’ ability to maximize individual rewards (Akata et al., 2025), while others investigate how strategic behavior changes when the model is prompted with contextual information or character-specific details (Lorè and Heydari, 2024), at times revealing altruistic tendencies (Phelps and Russell, 2024; Azaria, 2023) and human-like behavioral biases (Herr et al., 2024). In certain cases, LLMs have exhibited higher levels of cooperation (Brookins and DeBacker, 2023; Fontana et al., 2025) than predicted by standard GT models, suggesting the potential for reasoning beyond traditional payoff-maximization frameworks.

Another studied setting is ULTIMATUM GAME (UG, Harsanyi, 1961) and its variants, in which a proposer decides how to split a fixed sum of money with a recipient. The recipient can either accept the offer—resulting in the proposed split—or reject it, in which case neither player receives anything. In some variants, the recipient has no agency and must accept the offer. The game has long served as a benchmark for studying deviations from GT rationality, as players often demonstrate altruistic or fairness-oriented behavior that departs from strict payoff-maximization. In this regard, UG has been instrumental in probing economic decision-making and social preferences in both humans and artificial agents. Recent studies have examined LLM behavior in this context, testing their responses to abstract formulations of the game (Azaria, 2023; Johnson and Obradovich, 2023; Ross et al., 2024), as well as their adaptability to varying contextual prompts (Phelps and Russell, 2024; Filippas et al., 2024; Fan et al., 2024). In some cases, LLMs have demonstrated sensitivity to superficial details such as the names, genders, or titles of the players (Aher et al., 2023), revealing implicit biases and highlighting the importance of prompt phrasing in shaping strategic behavior.

Beyond these classical games, other studies have revealed additional tendencies in LLMs, including limited coordination capabilities (Akata et al., 2025), a preference for trustful strategies that yield favorable outcomes (Azaria, 2023), and occasional reasoning failures—even in simple or trivial game settings (Lorè and Heydari, 2024). Despite the variability in experimental designs and research objectives, a consistent trend emerges: the most advanced models, particularly those from the GPT (Achiam et al., 2023) and Llama (Touvron et al., 2023) families, exhibit a stronger grasp of game mechanics. These models are capable of interpreting personality traits specified in prompts (*e.g.*, selfishness, altruism) and can adjust their strategic and communicative behavior accordingly (Lorè and Heydari, 2024).

Finally, to support more systematic evaluations, several benchmarks have recently been introduced (Duan et al., 2024; Wang et al., 2024; Shapira et al., 2024; Huang et al., 2025). These benchmarks aim to improve the fairness and consistency of assessments by covering a wider range of game types, from one-shot scenarios to iterative and sequential interactions. Additionally, standardized experimental designs for testing strategic behavior in LLMs have been proposed (Hua et al., 2024), helping to align future work and enable more robust comparisons across studies.

6 CONCLUSION

In this work, we have examined the alignment between LLMs and humans in strategic behavior through a game-theoretic perspective. Specifically, we have evaluated LLM performance in three games where empirical findings diverge from theoretical predictions. Our analysis indicates that LLMs tend to imitate rational behavior, blending sometimes with human-like strategies. Moreover, LLMs appear mostly insensitive to numerical payoff-related hyperparameters in prompts, contrasting with typical human behavior.

Our findings provide a preliminary understanding of the alignment between LLMs and human strategic reasoning. Each game studied here deserves a more detailed analysis in future work, as the space of both game- and LLM-related hyperparameters is huge. As a next step, we aim to understand the observed misalignment of LLMs by leveraging insights from game theory literature, considering different behavioral biases (*e.g.*, altruism, risk aversion, etc.), and devising specific phenomenological models to better understand these behaviors.

REFERENCES

- 486
487
488 Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman,
489 Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report.
490 *arXiv preprint arXiv:2303.08774*, 2023.
- 491 Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. Using large language models to simulate
492 multiple humans and replicate human subject studies. In *International Conference on Machine*
493 *Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of*
494 *Machine Learning Research*, pages 337–371. PMLR, 2023.
- 495 Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz.
496 Playing repeated games with large language models. *Nature Human Behaviour*, 9(7):1380–1390,
497 July 2025.
- 498 James Andreoni and John H Miller. Rational cooperation in the finitely repeated prisoner’s dilemma:
499 Experimental evidence. *The economic journal*, 103(418):570–585, 1993.
- 500 Amos Azaria. Chatgpt: More human-like than computer-like, but not necessarily in a good way.
501 In Micah B. Goldwater, Florencia K. Anggoro, Brett K. Hayes, and Desmond C. Ong, editors,
502 *Proceedings of the 45th Annual Meeting of the Cognitive Science Society, CogSci 2023, Sydney,*
503 *NSW, Australia, July 26-29, 2023*. cognitivesciencesociety.org, 2023.
- 504 Kaushik Basu. The traveler’s dilemma: Paradoxes of rationality in game theory. *The American*
505 *Economic Review*, 84(2):391–395, 1994.
- 506 Kaushik Basu. The traveler’s dilemma. *Scientific American*, 296(6):90–95, 2007.
- 507 Kaushik Basu, Leonardo Becchetti, and Luca Stanca. Experiments with the Traveler’s Dilemma:
508 welfare, strategic choice and implicit collusion. *Soc. Choice Welf.*, 37(4):575–595, 2011.
- 509 Ken Binmore. Modeling rational players: Part I. *Economics & Philosophy*, 3(2):179–214, 1987.
- 510 Ken Binmore. *Natural justice*. Oxford university press, 2005.
- 511 Philip Brookins and Jason Matthew DeBacker. Playing games with gpt: What can we learn about a
512 large language model from canonical strategic games? *Available at SSRN 4493398*, 2023.
- 513 C Monica Capra, Jacob K Goeree, Rosario Gomez, and Charles A Holt. Anomalous behavior in a
514 traveler’s dilemma? *American Economic Review*, 89(3):678–690, 1999.
- 515 Kranti Chalamalasetti, Jana Götze, Sherzod Hakimov, Brielen Madureira, Philipp Sadler, and David
516 Schlangen. Clembench: Using game play to evaluate chat-optimized language models as con-
517 versational agents. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the*
518 *2023 Conference on Empirical Methods in Natural Language Processing*, pages 11174–11219,
519 Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.
520 emnlp-main.689. URL <https://aclanthology.org/2023.emnlp-main.689>.
- 521 Jinhao Duan, Renming Zhang, James Diffenderfer, Bhavya Kailkhura, Lichao Sun, Elias Stengel-
522 Eskin, Mohit Bansal, Tianlong Chen, and Kaidi Xu. Gtbench: Uncovering the strategic reasoning
523 limitations of llms via game-theoretic evaluations. *arXiv preprint arXiv:2402.12348*, 2024.
- 524 Francis Y Edgeworth. The pure theory of monopoly. *Papers relating to political economy*, 1(111):42,
525 1925.
- 526 Caoyun Fan, Jindou Chen, Yaohui Jin, and Hao He. Can large language models serve as rational
527 players in game theory? A systematic analysis. In *Thirty-Eighth AAAI Conference on Artificial*
528 *Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelli-*
529 *gence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI*
530 *2014, February 20-27, 2024, Vancouver, Canada*, pages 17960–17967. AAAI Press, 2024.
- 531 Apostolos Filippas, John J. Horton, and Benjamin S. Manning. Large language models as simulated
532 economic agents: What can we learn from homo silicus? In *Proceedings of the 25th ACM*
533 *Conference on Economics and Computation, EC ’24*, page 614–615, New York, NY, USA, 2024.
534 Association for Computing Machinery.

- 540 Nicoló Fontana, Francesco Pierri, and Luca Maria Aiello. Nicer than humans: How do large language
541 models behave in the prisoner’s dilemma? In *Proceedings of the Nineteenth International AAAI*
542 *Conference on Web and Social Media, June 23-26, 2025, Copenhagen, Denmark*, pages 522–535.
543 AAAI Press, 2025.
- 544 Takako Fujiwara-Greve. *Non-cooperative game theory*. Springer, 2015.
- 545
546 Yuling Gu, Oyvind Tafjord, Hyunwoo Kim, Jared Moore, Ronan Le Bras, Peter Clark, and Yejin
547 Choi. SimpleToM: Exposing the gap between explicit ToM inference and implicit ToM application
548 in LLMs, 2025. URL <https://arxiv.org/abs/2410.13648>.
- 549
550 Werner Güth, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum
551 bargaining. *Journal of economic behavior & organization*, 3(4):367–388, 1982.
- 552
553 John C Harsanyi. On the rationality postulates underlying the theory of cooperative games. *Journal*
554 *of Conflict Resolution*, 5(2):179–196, 1961.
- 555
556 Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob
557 Steinhardt. Measuring massive multitask language understanding. *Proceedings of the International*
Conference on Learning Representations (ICLR), 2021.
- 558
559 Nathan Herr, Fernando Acero, Roberta Raileanu, María Pérez-Ortiz, and Zhibin Li. Are large
560 language models strategic decision makers? A study of performance and bias in two-player
561 non-zero-sum games. *CoRR*, abs/2407.04467, 2024.
- 562
563 Ed Hopkins and Robert M Seymour. The stability of price dispersion under seller and consumer
564 learning. *International Economic Review*, 43(4):1157–1190, 2002.
- 565
566 Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin,
567 Lizhou Fan, Fei Sun, William Wang, et al. Game-theoretic llm: Agent workflow for negotiation
568 games. *arXiv preprint arXiv:2411.05990*, 2024.
- 569
570 Jen-tse Huang, Eric John Li, Man Ho Lam, Tian Liang, Wenxuan Wang, Youliang Yuan, Wenxiang
571 Jiao, Xing Wang, Zhaopeng Tu, and Michael Lyu. Competing large language models in multi-agent
572 gaming environments. In *The Thirteenth International Conference on Learning Representations*,
573 2025.
- 574
575 Zhijing Jin, Yuen Chen, Felix Leeb, Luigi Gresele, Ojasv Kamal, Zhiheng Lyu, Kevin Blin, Fernando
576 Gonzalez, Max Kleiman-Weiner, Mrinmaya Sachan, and Bernhard Schölkopf. CLadder: Assessing
577 causal reasoning in language models. In *NeurIPS*, 2023. URL [https://openreview.net/forum?](https://openreview.net/forum?id=e2wtjx0Yqu)
578 [id=e2wtjx0Yqu](https://openreview.net/forum?id=e2wtjx0Yqu).
- 579
580 Tim Johnson and Nick Obradovich. Evidence of behavior consistent with self-interest and altruism
581 in an artificially intelligent agent. *CoRR*, abs/2301.02330, 2023.
- 582
583 Steven Kuhn. Prisoner’s Dilemma. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford*
584 *Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2024 edition,
585 2024.
- 586
587 Nunzio Lorè and Babak Heydari. Strategic behavior of large language models and the role of game
588 structure versus contextual framing. *Scientific Reports*, 14(1):18490, 2024.
- 589
590 Richard D. McKelvey and Thomas R. Palfrey. An experimental study of the centipede game.
591 *Econometrica*, 60(4):803–836, 1992. ISSN 00129682, 14680262. URL [http://www.jstor.org/](http://www.jstor.org/stable/2951567)
592 [stable/2951567](http://www.jstor.org/stable/2951567).
- 593
594 Filippo Momentè, Alessandro Suglia, Mario Giulianelli, Ambra Ferrari, Alexander Koller, Oliver
595 Lemon, David Schlangen, Raquel Fernández, and Raffaella Bernardi. Triangulating LLM progress
596 through benchmarks, games, and cognitive tests, 2025. URL [https://arxiv.org/abs/2502.](https://arxiv.org/abs/2502.14359)
597 [14359](https://arxiv.org/abs/2502.14359).
- 598
599 John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the National Academy of*
600 *Sciences*, 36(1):48–49, 1950.

- 594 John F. Nash Jr. Non-cooperative games. In *Annals of Mathematics*, volume 54, pages 286–295.
595 Princeton University Press, 1951.
- 596
- 597 Steve Phelps and Yvan I. Russell. The machine psychology of cooperation: Can gpt models
598 operationalise prompts for altruism, cooperation, competitiveness and selfishness in economic
599 games?, 2024. URL <https://arxiv.org/abs/2305.07970>.
- 600 Robert W Rosenthal. Games of perfect information, predatory pricing and the chain-store paradox.
601 *Journal of Economic theory*, 25(1):92–100, 1981. doi: [https://doi.org/10.1016/0022-0531\(81\)](https://doi.org/10.1016/0022-0531(81)90018-1)
602 90018-1.
- 603
- 604 Jillian Ross, Yoon Kim, and Andrew W. Lo. Llm economicus? mapping the behavioral biases of llms
605 via utility theory. *arXiv preprint arXiv:*, 2024. accepted at COLM.
- 606 Eilam Shapira, Omer Madmon, Itamar Reinman, Samuel Joseph Amouyal, Roi Reichart, and
607 Moshe Tennenholtz. Glee: A unified framework and benchmark for language-based economic
608 environments. *arXiv preprint arXiv:2410.05254*, 2024.
- 609
- 610 Chandler Smith, Rakshit Trivedi, Jesse Clifton, Lewis Hammond, Akbir Khan, Sasha Vezhnevets,
611 John P. Agapiou, Edgar A. Duéñez-Guzmán, Jayd Matyas, Danny Karmon, Marwa Abdulhai,
612 Dylan Hadfield-Menell, Natasha Jaques, Joel Z. Leibo, Oliver Slumbers, Tim Baarslag, and Minsuk
613 Chang. The Concordia Contest: Advancing the cooperative intelligence of language agents, April
614 2024. URL <https://openreview.net/forum?id=dfefY1PSSw>.
- 615 Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam
616 Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska,
617 Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W.
618 Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda
619 Askeel, Amanda Dsouza, Ambrose Slone, Ameet Rahane, Anantharaman S. Iyer, Anders Johan
620 Andreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew M. Dai, Andrew
621 La, Andrew Kyle Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh
622 Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum,
623 Arul Menezes, Arun Kirubakaran, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia
624 Efrat, Aykut Erdem, Ayla Karakas, B. Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej
625 Bojanowski, Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno
626 Stein, Berk Ekmekci, Bill Yuchen Lin, Blake Howald, Bryan Orinion, Cameron Diao, Cameron
627 Dour, Catherine Stinson, Cedrick Argueta, Cesar Ferri, Chandan Singh, Charles Rathkopf, Chenlin
628 Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Christopher Waites, Christian Voigt, Christo-
629 pher D Manning, Christopher Potts, Cindy Ramirez, Clara E. Rivera, Clemencia Siro, Colin Raffel,
630 Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman,
631 Dan Roth, C. Daniel Freeman, Daniel Khashabi, Daniel Levy, Daniel Moseguí González, Danielle
632 Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David
633 Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Deniz
634 Yuret, Derek Chen, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho
635 Mollo, Diyi Yang, Dong-Ho Lee, Dylan Schrader, Ekaterina Shutova, Ekin Dogus Cubuk, Elad
636 Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele Rodolà,
637 Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan A Chi, Ethan Dyer, Ethan
638 Jerzak, Ethan Kim, Eunice Engefu Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar,
639 Fernando Martínez-Plumed, Francesca Happé, Francois Chollet, Frieda Rong, Gaurav Mishra,
640 Genta Indra Winata, Gerard de Melo, Germán Kruszewski, Giambattista Parascandolo, Giorgio
641 Mariani, Gloria Xinyue Wang, Gonzalo Jaimovitch-Lopez, Gregor Betz, Guy Gur-Ari, Hana
642 Galijasevic, Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar,
643 Henry Francis Anthony Shevlin, Hinrich Schuetze, Hiromu Yakura, Hongming Zhang, Hugh Mee
644 Wong, Ian Ng, Isaac Noble, Jaap Jumelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon
645 Lee, Jaime Fernández Fisac, James B Simon, James Koppel, James Zheng, James Zou, Jan Kocon,
646 Jana Thompson, Janelle Wingfield, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason
647 Phang, Jason Wei, Jason Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy
Kim, Jeroen Taal, Jesse Engel, Jesujoba Alabi, Jiacheng Xu, Jiaming Song, Jillian Tang, Joan
Waweru, John Burden, John Miller, John U. Balis, Jonathan Batchelder, Jonathan Berant, Jörg
Frohberg, Jos Rozen, Jose Hernandez-Orallo, Joseph Boudeman, Joseph Guerr, Joseph Jones,
Joshua B. Tenenbaum, Joshua S. Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth,

- 648 Karthik Gopalakrishnan, Katerina Ignatyeva, Katja Markert, Kaustubh Dhole, Kevin Gimpel,
649 Kevin Omondi, Kory Wallace Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar,
650 Kyle McDonell, Kyle Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin,
651 Lidia Contreras-Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble,
652 Ludwig Schmidt, Luheng He, Luis Oliveros-Colón, Luke Metz, Lütfi Kerem Senel, Maarten
653 Bosma, Maarten Sap, Maartje Ter Hoeve, Maheen Farooqi, Manaal Faruqui, Mantas Mazeika,
654 Marco Baturan, Marco Marelli, Marco Maru, Maria Jose Ramirez-Quintana, Marie Tolkiehn,
655 Mario Giulianelli, Martha Lewis, Martin Potthast, Matthew L Leavitt, Matthias Hagen, Mátyás
656 Schubert, Medina Orduna Baitemirova, Melody Arnaud, Melvin McElrath, Michael Andrew
657 Yee, Michael Cohen, Michael Gu, Michael Ivanitskiy, Michael Starritt, Michael Strube, Michał
658 Śwędrowski, Michele Bevilacqua, Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimeo Xu, Mirac
659 Suzgun, Mitch Walker, Mo Tiwari, Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhddeh Gheini,
660 Mukund Varma T, Nanyun Peng, Nathan Andrew Chi, Nayeon Lee, Neta Gur-Ari Krakover,
661 Nicholas Cameron, Nicholas Roberts, Nick Doiron, Nicole Martinez, Nikita Nangia, Niklas
662 Deckers, Niklas Muennighoff, Nitish Shirish Keskar, Niveditha S. Iyer, Noah Constant, Noah
663 Fiedel, Nuan Wen, Oliver Zhang, Omar Agha, Omar Elbaghdadi, Omer Levy, Owain Evans,
664 Pablo Antonio Moreno Casares, Parth Doshi, Pascale Fung, Paul Pu Liang, Paul Vicol, Pegah
665 Alipoormolabashi, Peiyuan Liao, Percy Liang, Peter W Chang, Peter Eckersley, Phu Mon Htut,
666 Pinyu Hwang, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour, Priti Oli, Qiaozhu Mei, Qing
667 Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer Gabriel, Rahel Habacker, Ra-
668 mon Risco, Raphaël Millière, Rhythm Garg, Richard Barnes, Rif A. Saurous, Riku Arakawa,
669 Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman Sitelew, Ronan Le Bras,
670 Rosanne Liu, Rowan Jacobs, Rui Zhang, Russ Salakhutdinov, Ryan Andrew Chi, Seungjae Ryan
671 Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif M. Mohammad, Sajant Anand,
672 Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel R. Bowman, Samuel Stern
673 Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah A. Rous, Sarik Ghazarian, Sayan Ghosh, Sean
674 Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi, Shadi Ham-
675 dan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi, Shixiang Shane
676 Gu, Shubh Pachchigar, Shubham Toshniwal, Shyam Upadhyay, Shyamolima Shammie Debnath,
677 Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini, Soo-Hwan
678 Lee, Spencer Torene, Sriharsha Hatwar, Stanislas Dehaene, Stefan Divic, Stefano Ermon, Stella
679 Biderman, Stephanie Lin, Stephen Prasad, Steven Piantadosi, Stuart Shieber, Summer Misherggi,
680 Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li, Tao Yu, Tariq Ali, Tat-
681 sunori Hashimoto, Te-Lin Wu, Théo Desbordes, Theodore Rothschild, Thomas Phan, Tianle Wang,
682 Tiberius Nkinyili, Timo Schick, Timofei Kornev, Titus Tunduny, Tobias Gerstenberg, Trenton
683 Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg,
684 Victoria Nyamai, Vikas Raunak, Vinay Venkatesh Ramasesh, vinay uday prabhu, Vishakh Pad-
685 makumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang
686 Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz,
687 Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Sophie Hao, Yifu Chen, Yonatan Belinkov,
688 Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J. Wang, Zirui
689 Wang, and Ziyi Wu. *Beyond the imitation game: Quantifying and extrapolating the capabilities of
690 language models. Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL
691 <https://openreview.net/forum?id=uyTL5Bvosj>. Featured Certification.
- 689 Haoran Sun, Yusen Wu, Yukun Cheng, and Xu Chu. Game theory meets large language models: A
690 systematic survey. *arXiv preprint arXiv:2502.09053*, 2025.
- 692 Mirac Suzgun, Nathan Scales, Nathanael Schärli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung,
693 Aakanksha Chowdhery, Quoc Le, Ed Chi, Denny Zhou, and Jason Wei. Challenging BIG-bench
694 tasks and whether chain-of-thought can solve them. In Anna Rogers, Jordan Boyd-Graber, and
695 Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*,
696 pages 13003–13051, Toronto, Canada, July 2023. Association for Computational Linguistics. doi:
697 10.18653/v1/2023.findings-acl.824. URL [https://aclanthology.org/2023.findings-acl.
698 824](https://aclanthology.org/2023.findings-acl.824).
- 699 Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée
700 Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and
701 efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

- John Von Neumann. On the theory of games of strategy. *Mathematische Annalen*, 100:295–320, 1928.
- Mark Walker and John Wooders. Minimax play at wimbledon. *American Economic Review*, 91(5): 1521–1538, 2001.
- Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. Superglue: A stickier benchmark for general-purpose language understanding systems. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019a. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/4496bf24afe7fab6f046bf4923da8de6-Paper.pdf.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *International Conference on Learning Representations*, 2019b. URL <https://openreview.net/forum?id=rJ4km2R5t7>.
- Haochuan Wang, Xiachong Feng, Lei Li, Zhanyue Qin, Dianbo Sui, and Lingpeng Kong. Tmgbench: A systematic game benchmark for evaluating strategic reasoning abilities of llms. *arXiv preprint arXiv:2410.10479*, 2024.
- Zhijian Wang, Bin Xu, and Hai-Jun Zhou. Social cycling and conditional responses in the rock-paper-scissors game. *Scientific reports*, 4(1):5830, 2014.
- Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.
- Bin Xu, Hai-Jun Zhou, and Zhijian Wang. Cycle frequency in standard rock–paper–scissors games: evidence from experimental economics. *Physica A: Statistical Mechanics and its Applications*, 392(20):4997–5005, 2013.
- Shunyu Yao, Noah Shinn, Pedram Razavi, and Karthik Narasimhan. τ -bench: A benchmark for tool-agent-user interaction in real-world domains, 2024. URL <https://arxiv.org/abs/2406.12045>.
- Andy K. Zhang, Neil Perry, Riya Dulepet, Joey Ji, Celeste Menders, Justin W. Lin, Eliot Jones, Gashon Hussein, Samantha Liu, Donovan Jasper, Pura Peetathawatchai, Ari Glenn, Vikram Sivashankar, Daniel Zamoshchin, Leo Glikbarg, Derek Askaryar, Mike Yang, Teddy Zhang, Rishi Alluri, Nathan Tran, Rinnara Sangpisit, Polycarpus Yiorkadjis, Kenny Osele, Gautham Raghupathi, Dan Boneh, Daniel E. Ho, and Percy Liang. Cybench: A framework for evaluating cybersecurity capabilities and risks of language models, 2025. URL <https://arxiv.org/abs/2408.08926>.
- Huaxiu Steven Zheng, Swaroop Mishra, Hugh Zhang, Xinyun Chen, Minmin Chen, Azade Nova, Le Hou, Heng-Tze Cheng, Quoc V. Le, Ed H. Chi, and Denny Zhou. NATURAL PLAN: Benchmarking llms on natural language planning, 2024. URL <https://arxiv.org/abs/2406.04520>.
- Hai-Jun Zhou. The Rock–Paper–Scissors game. *Contemporary Physics*, 57(2):151–163, 2016.

A NOTIONS OF GAME THEORY

In this section, we provide a primer in Game Theory useful to understand the main concepts in our work.

Normal-Form Game (NFG). A normal-form game (such as Traveler’s Dilemma) models strategic interactions where all players make their moves simultaneously and independently. Formally, an NFG is defined as a triple $\mathcal{G} = (\mathcal{P}, \mathcal{A}, \mathcal{U})$, consisting of a set of players \mathcal{P} , their respective strategy sets \mathcal{A} , and payoff functions \mathcal{U} . The **players**, denoted $\mathcal{P} = \{1, \dots, M\}$, are the agents making decisions in the game. Each player i has an associated **strategy set** \mathcal{A}_i , which contains all possible strategies—either discrete or continuous—that the player can choose. All strategy sets are collected in $\mathcal{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_M\}$. Once each player selects a strategy, they receive a payoff determined by their **payoff function** $u_i : \times_{j=1}^M \mathcal{A}_j \rightarrow \mathbb{R}$, which depends on the strategy of all

756 players. Similarly, $\mathcal{U} = \{u_1, \dots, u_M\}$. A combination of strategies $\{a_1, \dots, a_M\}$, where $a_i \in \mathcal{A}_i$, is
 757 called a **strategy profile**, representing a complete specification of all players’ choices. A strategy
 758 profile $a^* = \{a_1^*, \dots, a_M^*\}$ is a **Nash Equilibrium (NE)** if no player can improve their payoff by
 759 unilaterally deviating. That is, for every player i and for all alternative strategies $a_i \in \mathcal{A}_i$, it holds
 760 that

$$761 \quad u_i(a^*) \geq u_i(\{a_1^*, \dots, a_i, \dots, a_M^*\}). \quad (1)$$

762 In such a state, each player’s strategy is optimal given the strategies of the other players.

763
 764 **Extensive-Form Game (EFG).** An extensive-form game (such Centipede Game) can be viewed
 765 as a generalization of an NFG, that allows players to take sequential actions across multiple rounds.
 766 While a full formalization of EFGs involves several additional concepts beyond the scope of this work,
 767 we provide an intuitive yet rigorous overview. The structure of an EFG is commonly represented as a
 768 directed tree $\mathcal{K} = (V, E, T)$. Each vertex $v \in V$ corresponds to a decision point assigned to a player,
 769 and the outgoing edges $E_v \subseteq E$ represent the available actions at that point. The game begins at the
 770 root of the tree and progresses along a path determined by the players’ choices, eventually reaching a
 771 terminal node $t \in T$. Each terminal node specifies the payoffs for the players based on the actions
 772 taken along the path. A player’s strategy in an EFG consists of choosing an action at each vertex in
 773 the tree where they are required to make a decision³. A complete strategy profile is the combination
 774 of strategies for all players. As in the normal-form setting, a Nash Equilibrium is a strategy profile
 775 where no player has an incentive to unilaterally deviate.

776 **Evolutionary Game (EG).** An evolutionary game (such Evolutionary Rock, Paper and Scissors) is
 777 a rather loose concept encompassing several possible definitions, often applied in GT-based biological
 778 evolution dynamics (*i.e.*, Evolutionary Game Theory Weibull (1997)). It models the interaction of
 779 a population of **individuals** playing a series of multiple NFGs. At each step of the sequence (*i.e.*,
 780 **generation**), the individuals are randomly grouped in pairs and play the same NFG with the same
 781 strategy set \mathcal{A} and payoff $u : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$. Based on the received payoff, they can adapt their behavior
 782 accordingly for the next generation⁴. The focus in an EG is on the evolution of the entire population
 783 summarized in each generation by a **mixed strategy**. Formally, a mixed strategy at generation g
 784 is a probability distribution $p^{(g)} = (p_1^{(g)}, \dots, p_{|\mathcal{A}|}^{(g)})$ over the shared strategy set, where each $p_a^{(g)}$
 785 represents the share of individuals in the population playing strategy $a \in \mathcal{A}$. A strategy a^* is said
 786 to be **Evolutionary Stable (ES)** if a population in the vicinity of a^* converges to it. Additionally,
 787 $\mathcal{U}(a^*, a^*)$ is a NE, therefore a^* is also informally denoted as the NE of the game.

788 B NOTIONS OF LANGUAGE MODELS

789 B.1 ALPHABETS AND STRINGS

790
 791
 792 An **alphabet** Σ is a non-empty, finite set of elements called **symbols**. A **string** σ over alphabet Σ is
 793 a finite sequence $\sigma = \sigma_1 \dots \sigma_N$ of symbols, $\sigma_1, \dots, \sigma_N \in \Sigma$. For any alphabet Σ , let Σ^* denote the
 794 set of all strings over Σ . Furthermore, let $|\sigma|$ denote the string’s length N . Given a string σ such that
 795 $|\sigma| \geq t$, let $\sigma_{<t}$ denote the string of the first $t-1$ characters of σ . We write $\sigma \preceq \sigma'$ if σ is a prefix
 796 of σ' , and use \succeq and \succ to refer to the relations \preceq and \prec with their respective arguments transposed.
 797 The relation \preceq defines a partial order on Σ^* . Finally, we denote EOS as a distinguished **end-of-string**
 798 **symbol**, which is not a member of the alphabet Σ .

800 C FROM STRINGS TO STRATEGIES

801
 802 While GT players are constrained to choose only the strategies prescribed by the game, an LLM
 803 output can span a much wider vocabulary. In the following, we provide a formalization of the
 804 relationship between strings generated by LLMs and strategy sets, used in our experiments.

805 Given an alphabet Σ , its Kleene closure Σ^* is, in short, the set of all possible finite-length strings
 806 composed with Σ . Let, then, $(\Sigma^*, 2^{\Sigma^*}, \mathbb{P}_{\Sigma^*})$ be a **probability space over strings**, with (a countably
 807

808 ³This is independent of the player actually reaching the vertex in the game.

809 ⁴Other variants are possible, with the individuals playing fixed strategies and adapting the population shares of
 individuals, instead of their personal behaviors.

infinite) sample space Σ^* , σ -algebra 2^{Σ^*} , and a probability measure over the strings \mathbb{P}_{Σ^*} . Similarly, we define a **probability space over strategies** $(\mathcal{A}, 2^{\mathcal{A}}, \mathbb{P}_{\mathcal{A}})$. Here, \mathcal{A} contains a player’s strategies a_1, \dots, a_k and a symbol ψ denoting an invalid (*i.e.*, not present in the game) strategy⁵.

Now, let the **encoding function** $e : \Sigma^* \rightarrow \mathcal{A}$ be a measurable function that maps a string to a strategy. The probability measure $\mathbb{P}_{\mathcal{A}}$ is then the **pushforward measure** induced by e , defined as:

$$\mathbb{P}_{\mathcal{A}}(E) = \mathbb{P}_{\Sigma^*}(e^{-1}(E)), \quad \forall E \in 2^{\Sigma^*}. \quad (2)$$

Without loss of generality, for $a \in \mathcal{A}$, we can express:

$$\mathbb{P}_{\mathcal{A}}(a) = \mathbb{P}_{\mathcal{A}}(\{a\}) = \sum_{\mathbf{w} \in \Sigma^*} \mathbb{1}\{e(\mathbf{w}) = a\} \mathbb{P}_{\Sigma^*}(\mathbf{w}) = \sum_{\mathbf{w} \in e^{-1}(\{a\})} \mathbb{P}_{\Sigma^*}(\mathbf{w}). \quad (3)$$

where $\mathbb{1}\{e(\mathbf{w}) = a\}$ is an indicator function returning 1 if $e(\mathbf{w}) = a$ and 0 otherwise.

As an example consider the ROCK, PAPER, SCISSORS game. Suppose Δ is the alphabet used by an LM and Δ^* is its Kleene closure. Let $\mathcal{A} = \{R, P, S, \psi\}$ and $e : \Delta^* \rightarrow \mathcal{A}$ an encoding function mapping strings to strategies (*e.g.*, through regular expression matching). Then e induces a partition on Δ^* . In particular, let S_R, S_P, S_S be the sets of strings corresponding to each valid strategy. For example, $S_R = e^{-1}(\{R\}) = \{\text{rock, Rock, ROCK}\}$.

The measure \mathbb{P}_{Σ} assigns probabilities to the strings in Δ^* . Then, the probability of selecting a valid strategy, *e.g.*, R, is given by summing the probabilities of all strings mapped to R under e :

$$\mathbb{P}_{\mathcal{A}}(R) = \sum_{\delta \in \Delta^*} \mathbb{1}\{\delta \in S_R\} \mathbb{P}_{\Sigma}(\delta) = \sum_{\delta \in e^{-1}(\{R\})} \mathbb{P}_{\Sigma}(\delta) \quad (4)$$

In our work, we use a simple encoding function based on regular expressions to identify strategy choices from model outputs. A string is mapped to one of the strategies R, P, or S if and only if it exactly matches the regular expression "I choose (rock|paper|scissors)", as specified in the system prompt describing the game. With appropriate prompt design and conditioning (*cf.* App. F for detailed prompt templates), we observe that the probability assigned to the invalid strategy $P_{\mathcal{A}}(\psi)$ —*i.e.*, strings that do not match any valid strategy—is negligibly low. We estimate this leakage using Monte Carlo sampling (2560 samples), and find it to be approximately 0 for RPS, 0.012 for CG, and 0.008 for TD, using LLaMa-3.2-3B. The leakage is even smaller using larger LLMs. Given the small entity of the phenomenon, we handle the rare invalid outputs using a simple heuristic: when an invalid strategy is selected, we resample from the language model output.

C.1 LANGUAGE MODELS AND PREFIX PROBABILITY

A **language model** p_{Σ} is a probability distribution over Σ^* , where Σ denotes an alphabet. It is often useful to represent the probability of a string as a product of single-symbol **conditional prefix probability** distributions, $\vec{p}_{\Gamma}^{\rightarrow}(\cdot \mid \sigma_{<t})$:

$$p_{\Sigma}(\sigma) = \vec{p}_{\Gamma}^{\rightarrow}(\text{EOS} \mid \sigma) \prod_{t=1}^{|\sigma|} \vec{p}_{\Gamma}^{\rightarrow}(\sigma_t \mid \sigma_{<t}) \quad (5)$$

Each conditional prefix probability $\vec{p}_{\Gamma}^{\rightarrow}(\sigma_t \mid \sigma_{<t})$ is a distribution over the set $\Sigma \cup \{\text{EOS}\}$ and is defined as

$$\vec{p}_{\Gamma}^{\rightarrow}(\sigma' \mid \sigma) \stackrel{\text{def}}{=} \mathbb{P}_{Y \sim p_{\Sigma}} [Y \succeq \sigma \sigma' \mid Y \succeq \sigma] = \frac{\vec{p}_{\Gamma}^{\rightarrow}(\sigma \sigma')}{\vec{p}_{\Gamma}^{\rightarrow}(\sigma)} \quad (6)$$

$$\vec{p}_{\Gamma}^{\rightarrow}(\text{EOS} \mid \sigma) \stackrel{\text{def}}{=} \frac{p_{\Sigma}(\sigma)}{\vec{p}_{\Gamma}^{\rightarrow}(\sigma)} \quad (7)$$

⁵In game theory, each player i can have its own probability space, constructed on its “extended” strategy set $\mathcal{A}_i = \mathcal{A}_i \cup \{\psi\}$ (*cf.* App. A). In the games under scrutiny in this work, the players are equivalent and have the same set of strategies, thus we omit the extra indices.

The conditional prefix probability is well-defined only when $\vec{p}_\Gamma(\sigma) > 0$, a standard caveat for conditional probabilities.⁶ In turn, the **prefix probability** $\vec{p}_\Gamma(\sigma)$ is the probability that Y has σ as a prefix:

$$\vec{p}_\Gamma(\sigma) \stackrel{\text{def}}{=} \mathbb{P}[Y \sim p_\Sigma] Y \succeq \sigma = \sum_{\sigma' \in \Sigma^*} \mathbb{1}\{\sigma' \succeq \sigma\} p_\Sigma(\sigma') \quad (8)$$

This formulation underlies most modern autoregressive language models, where each conditional prefix probability is produced by a learned parametric model.

D EXPERIMENTAL SETTINGS

D.1 PRETRAINED LANGUAGE MODELS

In our experiments, we use pretrained language models from Huggingface, namely Llama-3.2-3B-Instruct⁷, Llama-3.1-8B-Instruct⁸, Llama-4-Scout⁹, OLMo-2-0325-32B-Instruct¹⁰, OLMo-2-1124-13B-Instruct¹¹, Gemma-3-12B-it¹², Gemma-3-27B-it¹³, and DeepSeek-V3-0324¹⁴. In our preliminary experiments, we found that varying the sampling temperature ([0.5, 2.0]) only marginally changes the strategy distribution of LLMs. Increasing the temperature beyond 2.0 resulted in significantly more invalid strategies. Thus, we use a default temperature of 1.0.

D.2 COMPUTE RESOURCES

All our experiments are conducted with 4 NVIDIA A100 80GB GPUs. For the DeepSeek-V3 model, we use Together API¹⁵ to obtain the responses. For all the experiments, we estimate a total of roughly one week of compute time.

E DETAILED EXPERIMENTAL RESULTS

E.1 ROCK, PAPER, SCISSORS: ADDITIONAL RESULTS

⁶In practice, this holds for softmax-normalized LMs, that assign nonzero probability to all symbols in Σ^* .

⁷<https://huggingface.co/meta-llama/Llama-3.2-3B-Instruct>

⁸<https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct>

⁹<https://huggingface.co/meta-llama/Llama-4-Scout-17B-16E-Instruct>

¹⁰<https://huggingface.co/allenai/OLMo-2-0325-32B-Instruct>

¹¹<https://huggingface.co/allenai/OLMo-2-1124-13B-Instruct>

¹²<https://huggingface.co/google/gemma-3-12b-it>

¹³<https://huggingface.co/google/gemma-3-27b-it>

¹⁴<https://huggingface.co/deepseek-ai/DeepSeek-V3-0324>

¹⁵<https://www.together.ai/>

918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971

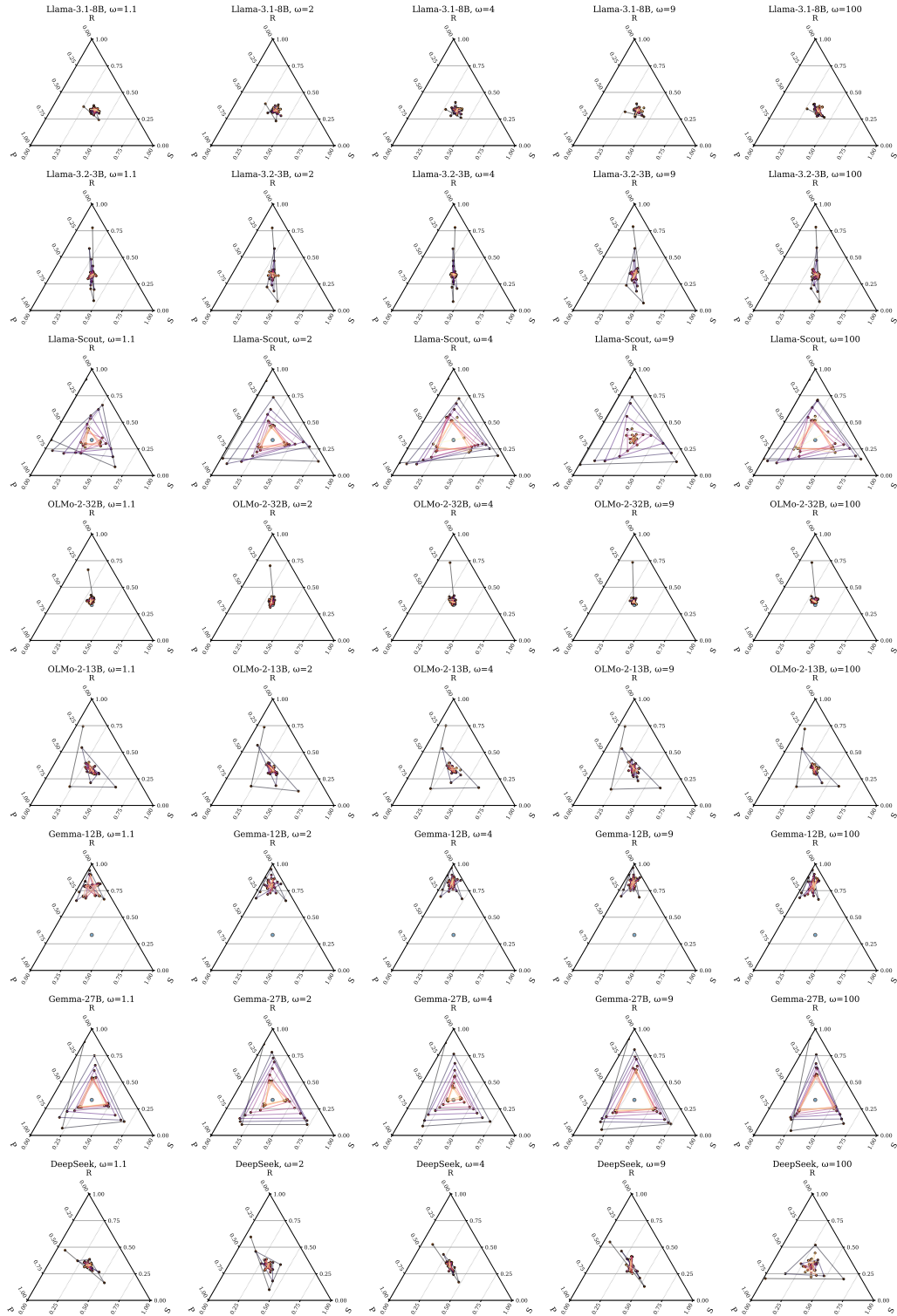


Figure 6: Cyclic trends in RPS across models. Columns (left→right): $\omega \in \{1.1, 2, 4, 9, 100\}$.

Table 3: **ROCK, PAPER, SCISSORS — Conditional response (extended)**. Extended version of Table 1, which reports the probability of replaying the previous round’s strategy. As in the main table, most LLMs—similar to humans (Wang et al., 2014, Fig. 2)—are more likely to repeat after a win (W_0) than after a loss or tie (L_0 , T_0), yet $W_0 < 1/3$ overall. Here we additionally report the *mean \pm standard deviation* across different verbalizations for each entry.

(a) $\omega \in \{1.1, 2, 4\}$

$p_0\%$	$\omega = 1.1$			$\omega = 2$			$\omega = 4$		
	W_0	T_0	L_0	W_0	T_0	L_0	W_0	T_0	L_0
Llama-3.2-3B	1.3 \pm 0.8	0.5 \pm 0.3	1.3 \pm 0.5	1.0 \pm 0.7	0.2 \pm 0.2	1.3 \pm 0.7	1.1 \pm 0.5	0.1 \pm 0.1	1.2 \pm 0.6
Llama-3.1-8B	14.9 \pm 1.8	9.5 \pm 1.6	11.2 \pm 1.4	15.2 \pm 2.0	9.1 \pm 1.4	11.0 \pm 1.6	15.7 \pm 2.8	8.7 \pm 1.6	10.9 \pm 1.4
Llama-Scout	4.3 \pm 2.4	2.2 \pm 3.0	0.1 \pm 0.2	2.4 \pm 2.3	0.7 \pm 1.7	0.1 \pm 0.2	12.7 \pm 30.8	0.5 \pm 1.0	0.1 \pm 0.1
OLMo-13B	1.1 \pm 0.6	0.7 \pm 0.5	0.5 \pm 0.5	1.0 \pm 0.8	0.4 \pm 0.3	0.4 \pm 0.4	1.2 \pm 0.8	0.4 \pm 0.3	0.3 \pm 0.2
OLMo-32B	11.2 \pm 5.5	9.6 \pm 6.2	5.6 \pm 3.7	9.0 \pm 3.8	6.1 \pm 3.3	3.3 \pm 2.5	10.1 \pm 4.4	7.6 \pm 6.4	4.1 \pm 2.3
Gemma-12B	42.7 \pm 28.8	67.7 \pm 43.6	35.4 \pm 22.5	42.2 \pm 28.9	69.0 \pm 43.0	46.3 \pm 28.2	45.9 \pm 33.6	69.9 \pm 43.1	38.4 \pm 30.7
Gemma-27B	54.9 \pm 26.4	17.1 \pm 11.7	14.3 \pm 10.6	46.6 \pm 28.3	15.8 \pm 12.9	20.2 \pm 14.6	53.1 \pm 24.7	18.4 \pm 12.8	23.8 \pm 15.0
DeepSeek	7.7 \pm 4.7	0.5 \pm 0.5	0.2 \pm 0.1	10.6 \pm 7.1	0.4 \pm 0.3	0.2 \pm 0.2	8.5 \pm 4.5	0.1 \pm 0.2	0.1 \pm 0.1

(b) $\omega \in \{9, 100\}$

$p_0\%$	$\omega = 9$			$\omega = 100$		
	W_0	T_0	L_0	W_0	T_0	L_0
Llama-3.2-3B	1.5 \pm 0.7	0.3 \pm 0.2	1.3 \pm 0.4	1.4 \pm 0.7	0.2 \pm 0.2	1.4 \pm 0.8
Llama-3.1-8B	16.1 \pm 1.5	9.8 \pm 1.2	11.4 \pm 1.2	15.0 \pm 1.3	9.3 \pm 0.9	11.1 \pm 1.0
Llama-Scout	3.0 \pm 1.9	0.4 \pm 0.9	0.1 \pm 0.2	1.8 \pm 1.8	0.6 \pm 1.8	0.1 \pm 0.2
OLMo-13B	1.1 \pm 0.6	0.5 \pm 0.4	0.3 \pm 0.2	1.5 \pm 0.9	0.8 \pm 0.5	0.6 \pm 0.4
OLMo-32B	10.4 \pm 4.5	8.4 \pm 5.1	3.3 \pm 2.0	12.6 \pm 5.7	9.2 \pm 6.1	4.7 \pm 2.6
Gemma-12B	51.5 \pm 35.8	71.9 \pm 41.5	53.3 \pm 33.6	52.1 \pm 28.9	72.1 \pm 39.6	56.6 \pm 30.4
Gemma-27B	38.8 \pm 28.6	13.2 \pm 11.0	17.2 \pm 14.6	40.8 \pm 29.7	14.3 \pm 12.1	15.9 \pm 14.2
DeepSeek	8.2 \pm 4.4	0.3 \pm 0.7	0.2 \pm 0.2	5.5 \pm 1.4	0.1 \pm 0.1	0.1 \pm 0.1

Table 4: **ROCK, PAPER, SCISSORS: Cyclic frequency**. $f_T = \frac{1}{2\pi T} \sum_{t=0}^T \theta(t)$, where $\theta(t)$ is the angle drawn by the trajectory after playing game t and T is the total number of played games (cf. Wang et al. (2014)). For each model and ω , we report the *mean \pm standard deviation* of f_T across verbalizations. Spearman’s coefficients ρ assess monotonic correlation with ω . Asterisks denote $p < 0.05$ from one-sided *permutation tests* of $H_0 : \mathbb{E}[f_T] = 0$ versus $H_1 : \mathbb{E}[f_T] < 0$.

$f_T \times 10^{-2}$	$\omega = 1.1$	$\omega = 2$	$\omega = 4$	$\omega = 9$	$\omega = 100$	ρ
Humans	3.1	2.7	3.1	2.2	1.8	-0.82
Llama-3.1-8B	-5.3 \pm 7.9	-4.2 \pm 5.7	-4.7 \pm 5.9	-4.9 \pm 4.3	-2.4 \pm 5.5	0.60
Llama-3.2-3B	1.7 \pm 16.4	0.8 \pm 11.0	4.7 \pm 9.3	7.7 \pm 11.0	4.4 \pm 12.1	0.60
Llama-Scout	-19.7 \pm 10.6	-15.5 \pm 13.3	-15.9 \pm 18.4	-15.9 \pm 11.8	-18.9 \pm 13.9	0.00
OLMo-13B	-8.7 \pm 8.6	-6.0 \pm 9.5	-8.6 \pm 8.0	-7.5 \pm 10.2	-5.0 \pm 8.4	0.70
OLMo-32B	0.7 \pm 4.1	1.3 \pm 8.3	-2.0 \pm 7.1	1.6 \pm 3.6	-1.9 \pm 5.0	-0.10
Gemma-12b	-3.0 \pm 10.3	-3.1 \pm 8.5	-2.4 \pm 7.4	-1.1 \pm 4.5	-1.1 \pm 4.3	0.90*
Gemma-27b	-13.0 \pm 13.2	-14.6 \pm 16.2	-11.8 \pm 11.1	-14.2 \pm 17.1	-10.5 \pm 19.3	0.50
DeepSeek	5.2 \pm 20.5	-6.1 \pm 20.1	2.7 \pm 18.7	1.4 \pm 22.0	-15.7 \pm 4.8	-0.70

* p -value < 0.05 (H_1 : mean < 0).

1026
 1027
 1028
 1029
 1030
 1031
 1032
 1033
 1034
 1035
 1036
 1037
 1038
 1039
 1040
 1041
 1042
 1043
 1044
 1045
 1046
 1047
 1048
 1049
 1050
 1051
 1052
 1053
 1054
 1055
 1056
 1057
 1058
 1059
 1060
 1061
 1062
 1063
 1064
 1065
 1066
 1067
 1068
 1069
 1070
 1071
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079

Table 5: **ROCK, PAPER, SCISSORS**– Average per-round payoff $\bar{\pi}$. Asterisks (*) indicate values significantly greater than the Nash equilibrium baseline $(a + 1)/3$ under a one-sided sign-flip permutation test ($p < 0.05$; alternative: model $>$ NE). Daggers (†) indicate values significantly less than the Nash equilibrium ($p < 0.05$; alternative: model $<$ NE). Wang et al. (2014) do not report exact human payoffs, but state they weakly exceed the NE strategy.

$\bar{\pi}$	$\omega = 1.1$	$\omega = 2$	$\omega = 4$	$\omega = 9$	$\omega = 100$
Humans	> 0.7	> 1	$> 1.\bar{6}$	$> 3.\bar{3}$	$> 33.\bar{6}$
DeepSeek	0.713*	0.986	1.625†	3.181†	31.966†
Llama-3.2-3B	0.724*	1.002	1.625†	3.182†	31.887†
Llama-3.1-8B	0.697	1.005	1.662	3.356	32.848†
Llama-Scout	0.777*	0.995	1.351†	2.824†	19.581†
OLMo-13B	0.724*	0.996	1.665	3.194†	31.705†
OLMo-32B	0.710*	1.006	1.662	3.290	32.320†
Gemma-12B	0.915*	1.002	1.199†	1.727†	11.958†
Gemma-27B	0.794*	1.004	1.490†	2.356†	21.820†
NASH EQ.	0.7	1	1. $\bar{6}$	3. $\bar{3}$	33. $\bar{6}$

* $p < 0.05$ ($H_1: \bar{\pi} > \text{NE}$)

† $p < 0.05$ ($H_1: \bar{\pi} < \text{NE}$)

E.2 CENTIPEDE GAME: ADDITIONAL RESULTS

Table 6: **CENTIPEDE GAME– Implied probabilities.** $p_t = e_t / \sum_{j=t}^n e_j$. For the (4, 1) payoff schedule we have $m_1 = 1$, $M_1 = 4$; when labeled *H.payoff* we use the high-stakes schedule (16, 4) with $m_1 = 4$, $M_1 = 16$. Cells report *mean ± standard deviation* across verbalizations; the value (G_t) is the mean number of sequences that reached round t . This table is an extended version of Table 2.

$p_t\%$ (G_t)	p_1	(G_1)	p_2	(G_2)	p_3	(G_3)	p_4	(G_4)	p_5	(G_5)	p_6	(G_6)	
4 Rounds	Humans	7	(281)	38	(261)	65	(161)	75	(57)				
	LLaMA-3.2-3B	30.1±14.6	(500)	21.8±12.0	(350)	26.1±15.0	(280)	29.7±17.5	(218)				
	LLaMA-3.1-8B	36.8±5.3	(500)	38.5±4.9	(316)	62.7±7.3	(195)	58.7±7.7	(74)				
	LLaMA-Scout	61.6±7.3	(500)	97.6±3.0	(192)	100.0±0.0	(4)	-	(0)				
	OLMo-13B	25.8±12.6	(500)	30.3±18.1	(371)	55.7±14.0	(269)	52.5±16.1	(130)				
	OLMo-2-32B	21.1±10.8	(500)	49.8±6.1	(394)	78.8±10.9	(200)	81.9±5.1	(42)				
	Gemma-12B	7.0±3.7	(500)	98.4±3.6	(465)	78.8±11.7	(7)	66.7±57.7	(2)				
	Gemma-27B	3.7±8.4	(500)	14.4±28.9	(482)	3.2±5.6	(422)	53.5±35.8	(413)				
DeepSeek	52.1±4.9	(500)	79.4±6.6	(239)	84.3±9.7	(50)	96.8±5.2	(8)					
4 Rounds (H. Payoff)	Humans	15	(100)	44	(85)	67	(48)	79	(16)				
	LLaMA-3.2-3B	24.6±10.9	(500)	22.2±12.1	(377)	26.2±12.6	(298)	28.0±16.5	(228)				
	LLaMA-3.1-8B	36.8±5.0	(500)	36.7±3.9	(316)	59.1±7.1	(201)	56.2±6.4	(83)				
	LLaMA-Scout	67.2±6.7	(500)	96.9±4.2	(164)	96.2±9.9	(5)	100.0±0.0	(0)				
	OLMo-13B	26.5±14.1	(500)	23.4±16.3	(368)	46.0±22.0	(292)	41.4±20.9	(176)				
	OLMo-2-32B	22.5±10.7	(500)	50.6±8.0	(388)	75.0±10.2	(194)	80.5±7.2	(48)				
	Gemma-12B	3.1±3.5	(500)	96.0±6.0	(485)	48.7±20.6	(19)	92.3±18.8	(11)				
	Gemma-27B	1.5±3.3	(500)	8.9±20.6	(493)	5.6±14.2	(452)	31.8±35.3	(440)				
DeepSeek	55.0±7.4	(500)	81.8±8.8	(225)	84.1±9.5	(44)	92.7±16.4	(8)					
6 Rounds	Humans	1	(281)	6	(279)	21	(261)	53	(205)	73	(97)	85	(26)
	LLaMA-3.2-3B	34.6±13.6	(500)	20.1±9.9	(327)	22.6±14.7	(266)	21.3±16.7	(215)	17.4±14.2	(181)	19.2±12.3	(158)
	LLaMA-3.1-8B	38.4±4.8	(500)	38.9±3.8	(308)	60.8±4.0	(189)	55.3±7.0	(74)	58.8±8.2	(33)	43.0±15.2	(14)
	LLaMA-Scout	60.9±7.3	(500)	94.3±9.7	(196)	96.7±9.9	(12)	100.0±0.0	(2)	-	(0)	-	(0)
	OLMo-13B	23.5±13.2	(500)	24.1±17.9	(382)	48.2±21.5	(299)	41.9±23.0	(174)	21.0±13.6	(124)	26.5±21.3	(108)
	OLMo-32B	17.9±11.6	(500)	44.4±9.9	(410)	79.1±6.8	(232)	77.4±4.5	(48)	77.8±15.1	(11)	87.0±18.2	(2)
	Gemma-12B	3.4±3.1	(500)	88.8±23.5	(483)	51.0±23.8	(55)	33.2±46.8	(40)	0.0±0.0	(6)	20.0±44.7	(6)
	Gemma-27B	0.5±1.0	(500)	1.9±4.6	(498)	0.0±0.0	(488)	0.3±0.8	(488)	0.0±0.0	(487)	17.9±24.0	(487)
DeepSeek	46.0±4.2	(500)	63.1±5.8	(270)	68.8±11.7	(100)	87.6±10.9	(33)	90.9±15.8	(5)	60.0±34.6	(1)	
6 Rounds (H. Payoff)	LLaMA-3.2-3B	30.1±15.1	(500)	19.0±10.2	(350)	23.1±18.4	(290)	22.6±18.5	(236)	15.2±10.4	(199)	19.3±16.1	(177)
	LLaMA-3.1-8B	37.5±6.7	(500)	34.2±3.8	(313)	59.4±7.1	(206)	54.5±10.1	(85)	60.9±14.3	(40)	35.8±18.9	(18)
	LLaMA-Scout	65.1±6.2	(500)	88.7±10.0	(175)	89.7±20.6	(21)	34.8±7.3	(4)	100.0±0.0	(3)	-	(0)
	OLMo-13B	25.1±16.0	(500)	18.7±16.3	(375)	34.3±21.2	(315)	23.5±18.4	(228)	13.2±9.3	(195)	13.7±8.5	(180)
	OLMo-32B	17.0±10.1	(500)	44.2±6.7	(415)	75.1±8.9	(234)	73.3±5.3	(58)	79.9±14.7	(15)	89.9±13.3	(4)
	Gemma-12B	1.0±1.3	(500)	87.2±21.0	(495)	26.5±18.5	(63)	28.4±39.4	(55)	2.6±5.3	(25)	0.0±0.0	(24)
	Gemma-27B	0.6±2.0	(500)	1.5±3.6	(497)	0.6±1.9	(490)	0.2±0.4	(487)	0.0±0.0	(487)	4.6±10.0	(487)
	DeepSeek	44.0±5.4	(500)	65.0±5.9	(280)	71.2±12.0	(99)	76.7±17.8	(29)	89.2±14.3	(8)	60.4±42.7	(2)

1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

Table 7: **CENTPEDE GAME– Cumulative outcome probabilities.** $F_t = \sum_{j=1}^t e_j$, where e_j is the proportion of games ending at round j . For the (4, 1) payoff schedule we have $m_1 = 1$, $M_1 = 4$; when labeled *H.payoff* we use the high-stakes schedule (16, 4) with $m_1 = 4$, $M_1 = 16$. Entries are *mean ± standard deviation* across verbalizations. Values typeset in **green** indicate an *increase* from rounds 1–5 to rounds 6–10; values in **red** indicate a *decrease*.

F_t (%)	F_1		F_2		F_3		F_4		F_5		F_6	
Rounds	1–5	6–10	1–5	6–10	1–5	6–10	1–5	6–10	1–5	6–10	1–5	6–10
Humans	6.2	8.1	36.5	49.3	72.4	87.5	92.4	97.8				
LLaMA-3.2-3B	31.0±14.1	29.2±15.8	43.2±18.2	44.7±18.3	55.3±19.3	57.7±20.8	63.8±20.8	69.6±18.6				
LLaMA-3.1-8B	34.6±5.0	38.9±6.3	60.2±5.9	61.7±6.8	83.8±4.4	86.4±6.1	92.2±3.8	95.2±2.2				
LLaMA-4-Scout	43.1±4.2	80.1±11.2	98.6±1.8	99.6±0.7	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0				
OLMo-13B	26.2±12.3	25.4±13.2	45.2±20.8	47.2±22.0	72.6±15.8	75.5±17.0	85.2±13.1	87.0±13.5				
OLMo-32B	17.2±7.3	25.0±15.0	51.3±8.7	68.7±11.3	86.9±7.2	96.2±3.2	97.0±2.7	99.5±0.5				
Gemma-12B	0.6±0.8	13.4±7.2	98.4±3.7	98.7±3.0	99.0±0.9	99.8±0.5	100.0±0.1	100.0±0.0				
Gemma-27B	1.5±3.3	5.8±13.4	11.2±23.1	20.1±39.5	11.8±24.4	23.0±39.3	55.6±36.7	64.5±42.1				
DeepSeek	41.5±4.5	62.8±5.7	82.0±6.7	97.9±1.4	96.7±2.4	100.0±0.0	99.8±0.3	100.0±0.0				
LLaMA-3.2-3B	26.9±10.2	22.3±12.1	39.6±15.9	41.1±17.0	53.8±16.3	55.1±19.7	63.2±17.4	66.6±20.3				
LLaMA-3.1-8B	34.4±5.4	39.2±5.4	58.6±5.6	61.0±6.2	81.6±5.0	85.0±5.5	90.3±3.5	94.9±2.5				
LLaMA-4-Scout	43.8±4.6	90.7±11.2	98.1±2.5	100.0±0.1	99.8±0.6	100.0±0.0	100.0±0.0	100.0±0.0				
OLMo-13B	26.6±13.6	26.4±14.8	39.9±20.1	43.4±23.9	64.1±22.4	65.5±26.9	74.8±21.1	74.9±26.3				
OLMo-32B	19.0±7.4	26.0±14.7	52.3±9.3	70.0±12.7	85.4±6.2	95.2±4.1	96.4±2.3	99.5±0.6				
Gemma-12B	0.4±0.9	5.7±6.4	96.0±6.0	96.4±5.3	96.6±5.6	98.9±2.1	99.8±0.6	100.0±0.1				
Gemma-27B	0.6±1.4	2.3±5.5	5.5±14.4	13.7±29.9	8.6±21.4	15.4±32.7	32.3±36.2	39.5±44.1				
DeepSeek	43.8±7.7	66.3±7.7	83.9±9.0	98.6±1.5	96.8±2.7	100.0±0.0	99.8±0.3	100.0±0.0				
Humans	0.0	1.5	5.5	8.9	22.7	31.7	55.8	75.8	88.9	92.7	97.9	99.3
LLaMA-3.2-3B	32.0±12.1	37.1±15.7	42.5±15.3	51.0±17.0	53.7±17.4	60.3±19.5	61.1±19.1	66.5±20.2	66.4±18.5	70.3±20.3	70.8±18.2	74.4±19.7
LLaMA-3.1-8B	32.9±5.7	44.0±4.7	60.8±5.3	63.7±5.3	83.4±3.4	86.8±3.5	91.3±2.7	95.4±1.7	96.1±1.7	98.2±0.9	97.4±1.6	99.2±0.4
LLaMA-4-Scout	42.1±5.3	79.6±10.6	97.0±6.3	98.2±3.4	99.2±2.4	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
OLMo-13B	22.4±12.3	24.6±14.5	37.3±20.4	43.1±23.8	63.2±23.5	67.3±25.8	74.6±24.2	75.6±26.0	79.2±21.5	77.6±24.6	82.3±20.2	79.4±24.4
OLMo-32B	14.4±7.5	21.4±16.1	43.6±10.1	63.7±16.0	83.6±6.3	97.3±1.6	95.9±1.9	99.7±0.3	99.1±0.8	100.0±0.1	99.8±0.2	100.0±0.0
Gemma-12B	0.1±0.2	6.7±6.2	88.8±23.5	89.2±22.5	91.0±21.6	93.1±19.4	98.4±3.0	99.3±1.4	98.4±3.0	99.3±1.4	98.4±3.0	99.4±1.4
Gemma-27B	0.2±0.4	0.8±1.6	1.2±2.9	3.4±7.8	1.2±2.9	3.4±7.8	1.8±4.2	3.4±7.8	1.8±4.2	3.4±7.8	18.3±21.8	21.7±26.2
DeepSeek	35.2±4.1	56.8±5.2	64.9±6.9	94.9±2.0	87.0±6.6	99.9±0.2	97.9±2.9	100.0±0.0	99.6±0.8	100.0±0.0	99.8±0.5	100.0±0.0
LLaMA-3.2-3B	29.1±13.5	31.1±17.6	38.9±17.0	45.3±19.5	50.2±20.4	55.3±23.1	58.1±22.5	62.1±23.8	63.2±21.8	65.9±23.3	67.4±21.9	70.3±22.9
LLaMA-3.1-8B	33.2±6.7	41.8±7.2	56.5±6.7	61.1±5.1	80.4±5.5	85.6±5.2	89.6±5.2	94.2±3.4	95.0±3.2	98.0±2.0	96.6±2.4	99.0±1.2
LLaMA-4-Scout	41.9±5.2	88.3±8.5	94.2±7.4	97.3±3.7	98.3±3.7	100.0±0.0	98.9±2.5	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
OLMo-13B	24.1±15.5	26.0±16.6	35.3±22.0	38.6±24.8	53.2±26.1	55.8±29.4	59.9±28.1	62.2±31.9	63.6±28.7	64.2±31.8	68.6±26.8	66.2±31.1
OLMo-2-32B	14.3±7.9	19.7±13.3	42.2±8.1	64.2±12.4	80.6±7.0	96.4±2.7	94.3±2.5	99.6±0.3	98.6±1.3	100.0±0.1	99.8±0.3	100.0±0.0
Gemma-12B	0.4±0.9	1.5±1.9	87.2±20.9	87.4±20.4	87.7±20.9	90.2±19.0	94.3±9.1	95.7±8.0	94.4±9.0	95.9±7.7	94.4±9.0	95.9±7.7
Gemma-27B	0.4±1.3	0.9±2.8	1.5±4.0	2.6±6.7	2.0±5.5	3.1±8.4	2.2±6.1	3.1±8.4	2.2±6.1	3.1±8.4	5.9±13.3	7.1±16.2
DeepSeek	31.0±5.2	57.0±6.4	64.0±7.7	96.4±2.4	88.6±5.3	100.0±0.1	96.8±3.3	100.0±0.0	99.3±1.2	100.0±0.0	99.8±0.4	100.0±0.0

Table 8: **CENTPEDE GAME– Expected payoffs of players.** Asterisks mark one-sided sign-flip permutation tests vs. humans.

$\bar{\pi}_i$	4 Rounds		4 Rounds (H. P.)		6 Rounds		6 Rounds (H. P.)	
	$\bar{\pi}_1$	$\bar{\pi}_2$	$\bar{\pi}_1$	$\bar{\pi}_2$	$\bar{\pi}_1$	$\bar{\pi}_2$	$\bar{\pi}_1$	$\bar{\pi}_2$
Humans	11.38	10.08	37.68	36.12	28.72	28.67	-	-
LLaMA-3. 2-3B	25.61*	10.51	107.36*	44.14	78.16*	27.65	350.88	120.70
LLaMA-3. 1-8B	10.55	7.03	44.67	29.23	13.51	8.97	61.97	38.81
LLaMA-Scout	3.35	3.65	13.89	13.13	3.52	3.75	16.58	13.68
OLMo-13B	15.76	9.08	87.79*	38.63	58.00	21.65	363.68	119.25
OLMo-32B	8.31	7.02	33.01	29.06	9.33	7.70	38.70	33.59
Gemma-12B	2.33	7.53	9.94	32.96	5.71	9.96	59.97	48.15
Gemma-27B	29.66*	21.10*	174.20*	74.67*	210.46*	73.76*	962.93	259.75
DeepSeek	4.36	4.40	16.95	16.96	6.00	5.82	24.84	23.23
NASH EQ.	4.00	1.00	16.00	4.00	4.00	1.00	16.00	4.00

p -value < 0.05 ($H_1: \bar{\pi}_i^{(LLM)} > \bar{\pi}_i^{(human)}$)

E.3 TRAVELER’S DILEMMA: ADDITIONAL RESULTS

Table 9: **TRAVELER’S DILEMMA: Average claims (symmetric r).** For each bonus/penalty $r \in \{5, 10, 20, 25, 50, 80\}$, columns “1” and “8–10” report the average claim in round 1 and in the last three rounds of the session, respectively. Entries are *mean \pm standard deviation* across verbalizations. Values typeset in **green** indicate an *increase* from round 1 to rounds 8–10; values in **red** indicate a *decrease*. Human results are shown for reference. NASH EQ. and PARETO EFF. provide theoretical benchmarks.

	$r = 5$		$r = 10$		$r = 20$		$r = 25$		$r = 50$		$r = 80$	
	1	8-10	1	8-10	1	8-10	1	8-10	1	8-10	1	8-10
Humans	180	195	177	186	131	119	150	138	155	85	120	81
LLaMA-3. 2-3B	146 \pm 33	169 \pm 28	148 \pm 29	168 \pm 26	144 \pm 25	166 \pm 28	147 \pm 28	165 \pm 29	148 \pm 24	163 \pm 29	148 \pm 26	165 \pm 29
LLaMA-3. 1-8B	168 \pm 32	179 \pm 27	168 \pm 30	181 \pm 28	164 \pm 30	174 \pm 31	169 \pm 30	177 \pm 27	172 \pm 31	178 \pm 30	184 \pm 25	174 \pm 33
LLaMA-Scout	116 \pm 32	198 \pm 9	148 \pm 7	176 \pm 19	130 \pm 14	187 \pm 23	88 \pm 20	199 \pm 11	149 \pm 5	200 \pm 0	80 \pm 0	165 \pm 54
OLMo-13B	108 \pm 31	170 \pm 29	113 \pm 29	174 \pm 28	109 \pm 25	168 \pm 29	111 \pm 23	170 \pm 28	119 \pm 26	171 \pm 23	130 \pm 32	168 \pm 27
OLMo-32B	140 \pm 44	167 \pm 36	142 \pm 44	164 \pm 35	144 \pm 40	169 \pm 30	144 \pm 45	174 \pm 32	136 \pm 44	163 \pm 34	147 \pm 36	159 \pm 31
Gemma-12B	96 \pm 12	97 \pm 12	101 \pm 5	108 \pm 17	96 \pm 20	99 \pm 23	103 \pm 11	109 \pm 17	117 \pm 15	122 \pm 20	101 \pm 11	104 \pm 16
Gemma-27B	80 \pm 2	80 \pm 2	149 \pm 54	144 \pm 53	100 \pm 1	100 \pm 3	100 \pm 1	101 \pm 3	120 \pm 14	137 \pm 32	121 \pm 8	121 \pm 8
DeepSeek	154 \pm 58	117 \pm 55	150 \pm 59	108 \pm 51	129 \pm 57	101 \pm 43	154 \pm 56	126 \pm 56	110 \pm 43	86 \pm 24	130 \pm 54	89 \pm 28
NASH EQ.	80	80	80	80	80	80	80	80	80	80	80	80
PARETO EFF.	200	200	200	200	200	200	200	200	200	200	200	200

Table 10: **TRAVELER’S DILEMMA– Average claims (asymmetric r)**. Each column r_1 - r_2 reports player 1’s average claim. Symmetric columns come from baseline runs with $r=10$ and $r=80$, pooling both players. Bold marks the larger mean within each pair.

	Asymmetric		Symmetric	
	10–80	80–10	10–10	80–80
Humans	149.1	112.3	174.4	115.6
LLaMA-3.2-3B	150.3*	137.9	147.8	147.6
LLaMA-3.1-8B	169.4	181.2	168.2	183.9
LLaMA-Scout	83.4	84.1	148.4	80.0
OLMo-13B	122.2 [†]	131.8	112.5	129.8
OLMo-32B	154.2[†]	148.1	141.8	146.6
Gemma-12B	99.5	104.0	100.6	101.0
Gemma-27B	183.8*[†]	131.6	149.2	121.2
DeepSeek	160.1	144.2	150.2	130.1
NASH EQ.	80.0	80.0	80.0	80.0
PARETO EFF.	200.0	200.0	200.0	200.0

* p -value < 0.05 (H_1 : 10-80 $>$ 80-10)

[†] p -value < 0.05 (H_1 : 10-80 $>$ 10-10)

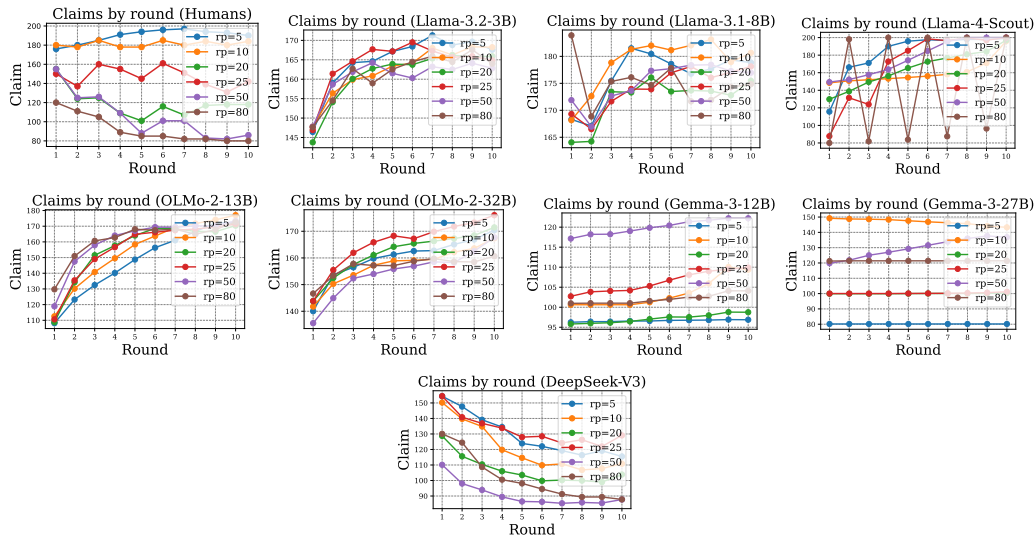


Figure 7: **TRAVELER’S DILEMMA– Claims by round (extended)**. Average claim across rounds for human players (top left, (Capra et al., 1999)) and for each LLM, under different bonus/penalty values $r_p \in \{5, 10, 20, 25, 50, 80\}$. Curves show how claims evolve within a 10-round session. Similar to humans, many models display systematic trends (increasing or decreasing) as the game progresses, while others remain relatively flat. This figure extends Fig. 7 by showing results for all models.

F PROMPT TEMPLATES

F.1 ROCK, PAPER, SCISSORS

```

1 {
2   "system": (
3     "You are a player in a game that lasts for {round_num} repeated rounds. "
4     "You will interact with a different player at each round. "
5     "At each round, you must emit exactly one symbol from the set: {rock}, {paper},
6     ↪ {scissors}. "
7     "Rules: {rock} defeats {scissors}, {scissors} defeats {paper}, and {paper}
8     ↪ defeats {rock}. "
9     "If your symbol defeats the other player’s, you receive {a:.1f} points, the other
10    ↪ player receives 0. "

```

```

1296 8      "If you are defeated, you receive 0 and the other player receives {a:.1f}. "
1297 9      "If symbols are identical, both players receive 1 point. "
1298 10     "Output constraint: reply EXACTLY 'I choose <symbol>', with <symbol> one of
1299     ↪ {rock}, {paper}, {scissors}. "
1300 11     "Your objective is to maximize your payoff at each round and overall. "
1301 12     )
1302 13 }
1303 14 {
1304 15     "user": (
1305 16         "Round {round_id}/{n_rounds}. Which option do you choose?"
1306 17     )
1307 18 }
1308 19 {
1309 20     "assistant": "I choose {option}"
1310 21 }
1311 22 {
1312 23     "user": (
1313 24         "In round {round_id}, you chose {self_option}, the other player chose
1314 25         ↪ {other_option}. "
1315 26         "Thus, you receive {self_gain} and the other player receives {other_gain}."
1316 27     )
1317 28 }
1318 29 }
1319 30 }
1320 31 }
1321 32 }
1322 33 }
1323 34 }
1324 35 }
1325 36 }
1326 37 }
1327 38 }
1328 39 }
1329 40 }
1330 41 }
1331 42 }
1332 43 }
1333 44 }
1334 45 }
1335 46 }
1336 47 }
1337 48 }
1338 49 }
1339 50 }
1340 51 }
1341 52 }
1342 53 }
1343 54 }
1344 55 }
1345 56 }
1346 57 }
1347 58 }
1348 59 }
1349 60 }
1350 61 }
1351 62 }
1352 63 }
1353 64 }
1354 65 }
1355 66 }
1356 67 }
1357 68 }
1358 69 }
1359 70 }
1360 71 }
1361 72 }
1362 73 }
1363 74 }
1364 75 }
1365 76 }
1366 77 }
1367 78 }
1368 79 }
1369 80 }
1370 81 }
1371 82 }
1372 83 }
1373 84 }
1374 85 }
1375 86 }
1376 87 }
1377 88 }
1378 89 }
1379 90 }
1380 91 }
1381 92 }
1382 93 }
1383 94 }
1384 95 }
1385 96 }
1386 97 }
1387 98 }
1388 99 }
1389 100 }
1390 101 }
1391 102 }
1392 103 }
1393 104 }
1394 105 }
1395 106 }
1396 107 }
1397 108 }
1398 109 }
1399 110 }
1400 111 }
1401 112 }
1402 113 }
1403 114 }
1404 115 }
1405 116 }
1406 117 }
1407 118 }
1408 119 }
1409 120 }
1410 121 }
1411 122 }
1412 123 }
1413 124 }
1414 125 }
1415 126 }
1416 127 }
1417 128 }
1418 129 }
1419 130 }
1420 131 }
1421 132 }
1422 133 }
1423 134 }
1424 135 }
1425 136 }
1426 137 }
1427 138 }
1428 139 }
1429 140 }
1430 141 }
1431 142 }
1432 143 }
1433 144 }
1434 145 }
1435 146 }
1436 147 }
1437 148 }
1438 149 }
1439 150 }
1440 151 }
1441 152 }
1442 153 }
1443 154 }
1444 155 }
1445 156 }
1446 157 }
1447 158 }
1448 159 }
1449 160 }
1450 161 }
1451 162 }
1452 163 }
1453 164 }
1454 165 }
1455 166 }
1456 167 }
1457 168 }
1458 169 }
1459 170 }
1460 171 }
1461 172 }
1462 173 }
1463 174 }
1464 175 }
1465 176 }
1466 177 }
1467 178 }
1468 179 }
1469 180 }
1470 181 }
1471 182 }
1472 183 }
1473 184 }
1474 185 }
1475 186 }
1476 187 }
1477 188 }
1478 189 }
1479 190 }
1480 191 }
1481 192 }
1482 193 }
1483 194 }
1484 195 }
1485 196 }
1486 197 }
1487 198 }
1488 199 }
1489 200 }
1490 201 }
1491 202 }
1492 203 }
1493 204 }
1494 205 }
1495 206 }
1496 207 }
1497 208 }
1498 209 }
1499 210 }
1500 211 }
1501 212 }
1502 213 }
1503 214 }
1504 215 }
1505 216 }
1506 217 }
1507 218 }
1508 219 }
1509 220 }
1510 221 }
1511 222 }
1512 223 }
1513 224 }
1514 225 }
1515 226 }
1516 227 }
1517 228 }
1518 229 }
1519 230 }
1520 231 }
1521 232 }
1522 233 }
1523 234 }
1524 235 }
1525 236 }
1526 237 }
1527 238 }
1528 239 }
1529 240 }
1530 241 }
1531 242 }
1532 243 }
1533 244 }
1534 245 }
1535 246 }
1536 247 }
1537 248 }
1538 249 }
1539 250 }
1540 251 }
1541 252 }
1542 253 }
1543 254 }
1544 255 }
1545 256 }
1546 257 }
1547 258 }
1548 259 }
1549 260 }
1550 261 }
1551 262 }
1552 263 }
1553 264 }
1554 265 }
1555 266 }
1556 267 }
1557 268 }
1558 269 }
1559 270 }
1560 271 }
1561 272 }
1562 273 }
1563 274 }
1564 275 }
1565 276 }
1566 277 }
1567 278 }
1568 279 }
1569 280 }
1570 281 }
1571 282 }
1572 283 }
1573 284 }
1574 285 }
1575 286 }
1576 287 }
1577 288 }
1578 289 }
1579 290 }
1580 291 }
1581 292 }
1582 293 }
1583 294 }
1584 295 }
1585 296 }
1586 297 }
1587 298 }
1588 299 }
1589 300 }
1590 301 }
1591 302 }
1592 303 }
1593 304 }
1594 305 }
1595 306 }
1596 307 }
1597 308 }
1598 309 }
1599 310 }
1600 311 }
1601 312 }
1602 313 }
1603 314 }
1604 315 }
1605 316 }
1606 317 }
1607 318 }
1608 319 }
1609 320 }
1610 321 }
1611 322 }
1612 323 }
1613 324 }
1614 325 }
1615 326 }
1616 327 }
1617 328 }
1618 329 }
1619 330 }
1620 331 }
1621 332 }
1622 333 }
1623 334 }
1624 335 }
1625 336 }
1626 337 }
1627 338 }
1628 339 }
1629 340 }
1630 341 }
1631 342 }
1632 343 }
1633 344 }
1634 345 }
1635 346 }
1636 347 }
1637 348 }
1638 349 }
1639 350 }
1640 351 }
1641 352 }
1642 353 }
1643 354 }
1644 355 }
1645 356 }
1646 357 }
1647 358 }
1648 359 }
1649 360 }
1650 361 }
1651 362 }
1652 363 }
1653 364 }
1654 365 }
1655 366 }
1656 367 }
1657 368 }
1658 369 }
1659 370 }
1660 371 }
1661 372 }
1662 373 }
1663 374 }
1664 375 }
1665 376 }
1666 377 }
1667 378 }
1668 379 }
1669 380 }
1670 381 }
1671 382 }
1672 383 }
1673 384 }
1674 385 }
1675 386 }
1676 387 }
1677 388 }
1678 389 }
1679 390 }
1680 391 }
1681 392 }
1682 393 }
1683 394 }
1684 395 }
1685 396 }
1686 397 }
1687 398 }
1688 399 }
1689 400 }
1690 401 }
1691 402 }
1692 403 }
1693 404 }
1694 405 }
1695 406 }
1696 407 }
1697 408 }
1698 409 }
1699 410 }
1700 411 }
1701 412 }
1702 413 }
1703 414 }
1704 415 }
1705 416 }
1706 417 }
1707 418 }
1708 419 }
1709 420 }
1710 421 }
1711 422 }
1712 423 }
1713 424 }
1714 425 }
1715 426 }
1716 427 }
1717 428 }
1718 429 }
1719 430 }
1720 431 }
1721 432 }
1722 433 }
1723 434 }
1724 435 }
1725 436 }
1726 437 }
1727 438 }
1728 439 }
1729 440 }
1730 441 }
1731 442 }
1732 443 }
1733 444 }
1734 445 }
1735 446 }
1736 447 }
1737 448 }
1738 449 }
1739 450 }
1740 451 }
1741 452 }
1742 453 }
1743 454 }
1744 455 }
1745 456 }
1746 457 }
1747 458 }
1748 459 }
1749 460 }
1750 461 }
1751 462 }
1752 463 }
1753 464 }
1754 465 }
1755 466 }
1756 467 }
1757 468 }
1758 469 }
1759 470 }
1760 471 }
1761 472 }
1762 473 }
1763 474 }
1764 475 }
1765 476 }
1766 477 }
1767 478 }
1768 479 }
1769 480 }
1770 481 }
1771 482 }
1772 483 }
1773 484 }
1774 485 }
1775 486 }
1776 487 }
1777 488 }
1778 489 }
1779 490 }
1780 491 }
1781 492 }
1782 493 }
1783 494 }
1784 495 }
1785 496 }
1786 497 }
1787 498 }
1788 499 }
1789 500 }
1790 501 }
1791 502 }
1792 503 }
1793 504 }
1794 505 }
1795 506 }
1796 507 }
1797 508 }
1798 509 }
1799 510 }
1800 511 }
1801 512 }
1802 513 }
1803 514 }
1804 515 }
1805 516 }
1806 517 }
1807 518 }
1808 519 }
1809 520 }
1810 521 }
1811 522 }
1812 523 }
1813 524 }
1814 525 }
1815 526 }
1816 527 }
1817 528 }
1818 529 }
1819 530 }
1820 531 }
1821 532 }
1822 533 }
1823 534 }
1824 535 }
1825 536 }
1826 537 }
1827 538 }
1828 539 }
1829 540 }
1830 541 }
1831 542 }
1832 543 }
1833 544 }
1834 545 }
1835 546 }
1836 547 }
1837 548 }
1838 549 }
1839 550 }
1840 551 }
1841 552 }
1842 553 }
1843 554 }
1844 555 }
1845 556 }
1846 557 }
1847 558 }
1848 559 }
1849 560 }
1850 561 }
1851 562 }
1852 563 }
1853 564 }
1854 565 }
1855 566 }
1856 567 }
1857 568 }
1858 569 }
1859 570 }
1860 571 }
1861 572 }
1862 573 }
1863 574 }
1864 575 }
1865 576 }
1866 577 }
1867 578 }
1868 579 }
1869 580 }
1870 581 }
1871 582 }
1872 583 }
1873 584 }
1874 585 }
1875 586 }
1876 587 }
1877 588 }
1878 589 }
1879 590 }
1880 591 }
1881 592 }
1882 593 }
1883 594 }
1884 595 }
1885 596 }
1886 597 }
1887 598 }
1888 599 }
1889 600 }
1890 601 }
1891 602 }
1892 603 }
1893 604 }
1894 605 }
1895 606 }
1896 607 }
1897 608 }
1898 609 }
1899 610 }
1900 611 }
1901 612 }
1902 613 }
1903 614 }
1904 615 }
1905 616 }
1906 617 }
1907 618 }
1908 619 }
1909 620 }
1910 621 }
1911 622 }
1912 623 }
1913 624 }
1914 625 }
1915 626 }
1916 627 }
1917 628 }
1918 629 }
1919 630 }
1920 631 }
1921 632 }
1922 633 }
1923 634 }
1924 635 }
1925 636 }
1926 637 }
1927 638 }
1928 639 }
1929 640 }
1930 641 }
1931 642 }
1932 643 }
1933 644 }
1934 645 }
1935 646 }
1936 647 }
1937 648 }
1938 649 }
1939 650 }
1940 651 }
1941 652 }
1942 653 }
1943 654 }
1944 655 }
1945 656 }
1946 657 }
1947 658 }
1948 659 }
1949 660 }
1950 661 }
1951 662 }
1952 663 }
1953 664 }
1954 665 }
1955 666 }
1956 667 }
1957 668 }
1958 669 }
1959 670 }
1960 671 }
1961 672 }
1962 673 }
1963 674 }
1964 675 }
1965 676 }
1966 677 }
1967 678 }
1968 679 }
1969 680 }
1970 681 }
1971 682 }
1972 683 }
1973 684 }
1974 685 }
1975 686 }
1976 687 }
1977 688 }
1978 689 }
1979 690 }
1980 691 }
1981 692 }
1982 693 }
1983 694 }
1984 695 }
1985 696 }
1986 697 }
1987 698 }
1988 699 }
1989 700 }
1990 701 }
1991 702 }
1992 703 }
1993 704 }
1994 705 }
1995 706 }
1996 707 }
1997 708 }
1998 709 }
1999 710 }
2000 711 }

```

```
1350 59 "You are a player in a game of {n_rounds} rounds. "
1351 60 "In each round, you and another player simultaneously select an integer k in [80,
1352 61 ↪ {max_value}]. "
1353 61 "Payoff rules: "
1354 62 "1. If both select the same k, each receives k. "
1355 63 "2. If your k < opponent's k: you receive k + {bp_value_1}, opponent receives k -
1356 64 ↪ {bp_value_2}. "
1357 64 "3. If your k > opponent's k: you receive opponent's k - {bp_value_1}, opponent
1358 65 ↪ receives opponent's k + {bp_value_2}. "
1359 66 "Your role: maximize your payoff each round and across the game. "
1360 66 "Output constraint: reply EXACTLY 'I choose <k>', where <k> is an integer in [80,
1361 67 ↪ {max_value}]. "
1362 67 )
1363 68 }
1364 69 {
1365 70 "user": (
1366 71 "Round {round_id}/{n_rounds}. What integer value do you select?"
1367 72 )
1368 73 }
1369 74 {
1370 75 "assistant": "I choose {option}"
1371 76 }
1372 77 {
1373 78 "user": (
1374 79 "In round {round_id}, you chose {self_option}, the other player chose
1375 80 ↪ {other_option}. "
1376 80 "Thus, you receive {self_gain}, the other player receives {other_gain}."
1377 81 )
1378 82 }
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
```