

Escaping the Nash Trap: Structural Estimation and Alignment of Strategic Reasoning in Large Language Models

Jiannan Xu^{1*} Yongkang Duan^{2*} Jane Yi Jiang³ Jiding Zhang^{4†}

¹University of Maryland

²Beijing Normal University

³The Ohio State University

⁴Arizona State University

Abstract

As large language models (LLMs) are increasingly deployed as decision-making agents in strategic environments, their performance depends critically on how they model human counterparts. Drawing on level- k thinking, we design normal-form games and develop a structural estimator that infers, from observed choices, the strategic reasoning type most consistent with an LLM’s behavior. Across models, we find a systematic mismatch: LLMs overwhelmingly behave as if their counterparts are fully rational NASH players and respond with Nash-oriented actions, whereas human participants in our online experiments exhibit substantial heterogeneity, ranging from RANDOM to NASH. This mismatch can help or hurt. An LLM may outsmart a boundedly rational human, but overestimating human sophistication can also create a *Nash equilibrium trap*, in which Nash play is not the payoff-maximizing response to observed human behavior. We design two prompting interventions to address this problem. The behaviorally informed intervention can substantially reduce Nash-oriented behavior, but it often induces untargeted deviations, including departures from Nash when equilibrium play remains appropriate. The welfare-aware intervention reduces Nash play less aggressively but yields higher empirical best-response rates and smaller welfare losses. Effective strategic alignment therefore requires not simply moving models away from Nash, but calibrating behavior to the relevant counterpart population and evaluating deviations against an explicit payoff or welfare objective.

1 Introduction

Large language model (LLM) agents are increasingly deployed in settings where they do not merely generate text but support decisions involving other agents. For example, they can assist with pricing in competitive markets (e.g., Wang et al. 2025), bidding in advertising auctions (e.g., Cai et al. 2025), hiring in labor markets (e.g., Lo et al. 2025), or supporting negotiations (e.g., Bianchi et al. 2024). In these settings, an agent’s payoff depends on both its own action, and its expectations about how a counterpart will respond. This raises a fundamental question: *what behavioral model of counterpart rationality best explains an LLM’s strategic choices, and when is that model well calibrated to human play?*

To study this question, we use the level- k framework from behavioral game theory (Stahl & Wilson, 1995). The level- k framework defines a set of strategic types (e.g., LEVEL-0, LEVEL-1, LEVEL-2, NASH (i.e., LEVEL- ∞)) where higher levels best respond to lower levels under a quantal-response model. We hence design twelve symmetric normal-form games grounded in inventory stocking, pricing competition, and advertising decisions, and test them on agents (LLMs or human subjects). For each agent, we estimate the action distribution implied by each strategic type, and use the agent’s observed choices across games to assign a maximum-a-posteriori (MAP) strategic type to that agent. These strategic type labels summarize behavioral consistency under the level- k quantal-response model.

*Equal contribution. Code is available at https://github.com/jiannan-xu/colm2026_nash_trap.

†Correspondence to: jiding.zhang@asu.edu.

Across models and conditions, LLMs choose Nash equilibrium actions far more often than humans, resulting in the NASH type being assigned to the majority of LLMs. In contrast, human behavior is substantially more heterogeneous. Participant-level MAP estimates classify only 12.5%–14.4% of condition records as NASH, with most assigned to LEVEL-1 or LEVEL-2. A complementary finite-mixture analysis in [Appendix F.4](#) likewise places most population mass on non-NASH types for human agents.

This mismatch is important: strategic choices determine the LLM’s payoff, the human counterpart’s payoff, and their joint payoff. In some games, Nash play allows the LLM to earn more against boundedly rational humans; in others, it is poorly calibrated to the observed human choice distribution. We say that a game exhibits a *Nash equilibrium trap* if the empirical best response to human choices differs from the unique Nash action; we refer to such a game as a *Nash-trap game*. While a Nash trap game does not imply that the Nash equilibrium is socially inefficient, it exhibits a mismatch between Nash play and the payoff-maximizing response to observed human behavior, and thus NASH-type LLMs may fail to respond optimally when facing human counterparts.

To address this challenge, we make four contributions:

1. We introduce a structural estimation framework for inferring latent strategic types from LLM and human choices in normal-form games. These types provide a behavioral characterization of strategic reasoning based on observed choices.
2. We document a robust action-level mismatch: across models and experimental conditions, 22 of 24 LLM estimates are classified as NASH, compared with only 13.4% of 186 human estimates. Although the precise distribution across lower-level types depends on the LEVEL-0 specification, human strategic behavior remains substantially more heterogeneous.
3. We show that the consequences of this mismatch are game-dependent. In some games, Nash play may increase an LLM’s payoff against boundedly rational humans. In Nash-trap games, however, it is not the empirical best response to observed human choices and therefore reduces the LLM’s own payoff.
4. We design behaviorally informed and welfare-aware prompting interventions to address this mismatch. Simply reducing Nash play often induces untargeted deviations that lower counterpart or joint payoff. Welfare-aware prompting produces higher empirical best-response rates and smaller welfare losses in Nash-trap games, although it is not uniformly superior across all objectives.

These findings have implications for deploying LLMs in strategic interactions involving human counterparts. The broader lesson is not to replace Nash reasoning universally, but to identify when Nash-oriented behavior is miscalibrated and evaluate alternative strategies against the payoff or welfare objectives relevant to the deployment context.

2 Literature review

2.1 Game theory and LLMs

Recently, there has been growing interest in studying LLMs’ behavior in game-theoretic settings. We refer readers to [Hua et al. \(2024\)](#) and [Sun et al. \(2025\)](#) for comprehensive reviews. Existing work largely falls into two categories. The first evaluates LLM performance in strategic environments, including matrix games (e.g., [Guo 2023](#); [Gandhi et al. 2023](#); [Xu et al. 2024](#); [Brookins & DeBacker 2024](#); [Fan et al. 2024](#); [Phelps & Russell 2025](#)) and repeated games (e.g., [Huang et al. 2024](#); [Akata et al. 2025](#); [Mao et al. 2025](#)). The second incorporates insights from behavioral game theory to explain deviations from Nash equilibrium and probe underlying reasoning patterns (e.g., [Wu et al. 2024](#); [Jia et al. 2025](#); [Zheng et al. 2025](#); [Pechon-Elkins & Chun 2026](#)).

However, prior work primarily focuses on observed actions or equilibrium outcomes, offering limited insight into the latent strategic reasoning processes that underlie these behaviors. In particular, there is little work that explicitly identifies and aligns LLMs’

strategic types with those of human decision makers. We address this gap by developing a structural estimation framework that (i) recovers latent strategic reasoning types from observed choices and (ii) enables direct comparison between LLMs and humans. We further introduce prompting interventions that encourage opponent-aware reasoning and evaluate when such behavioral shifts translate into improved performance against human play.

2.2 Level- k thinking and behavioral game theory

Level- k thinking is a foundational concept in behavioral game theory (e.g., Nagel 1995; Arad & Rubinstein 2012; Alaoui et al. 2020). It models strategic reasoning as a hierarchical process in which players iteratively best respond to beliefs about others’ behavior: level-0 players behave non-strategically, level-1 players best respond to level-0, level-2 players best respond to level-1, and so on. This framework provides a tractable way to capture boundedly rational reasoning in strategic environments.

Recent work has begun to apply this framework to evaluate and improve LLMs’ boundedly rational reasoning. For example, Jia et al. (2025) propose an evaluation framework for measuring LLMs’ level- k reasoning ability, while Zhang et al. (2025) introduce recursive prompting methods to induce higher-order beliefs. Despite these advances, there is limited empirical evidence on whether LLMs and humans exhibit systematically different distributions of strategic types. Moreover, how to design interventions such as prompting or fine-tuning to align LLMs’ strategic reasoning with human behavior remains largely unexplored. Our work contributes to this literature by structurally identifying strategic types and providing an intervention method to align them.

2.3 Human-centered LLMs

A growing body of research studies LLMs from a human-centered perspective, focusing on their ability to replicate or augment human cognition and decision-making at both individual and collective levels (e.g., Wang et al. 2024; Di Fede et al. 2025; Liu et al. 2025). Within this literature, reasoning is a central capability, motivating the development of benchmark datasets that evaluate LLM performance in complex decision tasks (e.g., Duan et al. 2024; Zhang et al. 2024).

However, such benchmarks primarily evaluate predictive accuracy or task performance, without identifying the underlying reasoning processes that drive model behavior. As a result, they provide limited insight into whether LLMs reason in ways that are comparable to humans. Our work complements this literature by moving from outcome-based evaluation to a structural behavioral analysis of strategic choice. By recovering latent strategic types from observed decisions, our framework enables a more disciplined comparison of LLM and human behavior and provides a foundation for aligning LLM agents with human decision-making in strategic settings.

3 Methodology

3.1 Overview

Our objective is to characterize the strategic type most consistent with observed choices of LLMs and human participants across a set of normal-form games. Since both humans and LLMs exhibit stochastic choice, we treat type assignment as a structural estimation problem rather than classifying behavior from a single action in a single game. Each strategic type induces a choice distribution over actions in each game. We estimate a common decision-precision parameter by maximum likelihood and compute posterior probabilities over types using the full pattern of choices across games. The resulting type estimates provide a behavioral characterization of strategic reasoning under a specified LEVEL-0 anchor and quantal-response choice model. We then evaluate whether behaviorally grounded prompting changes LLM action distributions and whether any change improves the LLM payoff, counterpart payoff, or joint payoff against observed human choices.

3.2 Experimental design

Game setup. We use $J = 12$ symmetric 3×3 normal-form games drawn from three business contexts: inventory stocking, pricing, and advertising. The design consists of six base games and six payoff-scaled variants that preserve the strategic structure while changing payoff magnitudes (Appendix A). For each game $j \in \{1, \dots, J\}$, let $A_j \in \mathbb{R}^{3 \times 3}$ denote the row player’s payoff matrix and $B_j = A_j^\top$ the column player’s payoff matrix.

Experimental implementation. We query each LLM for 30 independent responses per game under controlled prompts, yielding 360 nominal observations per model through API; invalid outputs are excluded after parsing. Human participants ($N = 186$) play the same 12 games once each in randomized order without feedback, so the benchmark captures one-shot decisions rather than learning.

Models, information conditions, and prompting approaches. We test six model families (gpt-5.2, claude-sonnet-4.5, gemini-3-flash-preview, deepseek-v3.2, qwen3-235b, and grok-4.3; Appendix B). Games appear either as abstract payoff matrices (*no context*) or in business scenarios preserving the same incentives (*contextualized*). Direct prompts request an action, whereas reasoning prompts request an explanation before the final action. Full templates and interventions are in Appendix C.

3.3 Descriptive evidence

Before conducting the structural estimation, we first summarize the extent of Nash-oriented behavior using the *Nash ratio*. For each game j , let a_j denote the observed action and n_j denote the unique pure-strategy Nash equilibrium action. Define

$$r_{\text{Nash}} = \frac{1}{J} \sum_{j=1}^J \mathbf{1}\{a_j = n_j\}, \quad (1)$$

which measures the fraction of games in which the observed action coincides with the Nash equilibrium action. For LLMs, we compute the Nash ratio for each repeated sample and report its average across samples. For human participants, we compute the ratio separately for each individual and report the corresponding aggregate statistics.

3.4 Structural estimation

We estimate the model separately for each agent under each experimental condition. For an LLM, the data consist of a $J \times 3$ count matrix $C = (c_{ja})$, where c_{ja} is the number of times action $a \in \{A, B, C\}$ is selected in game j across repeated samples. For a human participant, $c_{ja} \in \{0, 1\}$ and $\sum_{a=1}^3 c_{ja} = 1$. We assume that each agent belongs to one latent strategic type $t \in \mathcal{T} := \{0, 1, 2, \text{Nash}\}$, corresponding to RANDOM, LEVEL-1, LEVEL-2, and NASH reasoning. The estimator identifies the type-specific stochastic choice model that best explains the agent’s pattern of actions across the full game set. Accordingly, a MAP NASH label means that the observed cross-game choice pattern is best fit by the NASH type under our specification.

Given a belief q about the opponent’s mixed strategy, the row player’s logit quantal-response probabilities are

$$\text{QR}_{\text{row}}(A_j, q; \gamma)_a = \frac{\exp(\gamma[A_j q]_a)}{\sum_{a'=1}^3 \exp(\gamma[A_j q]_{a'})}, \quad (2)$$

where $\gamma > 0$, the decision precision parameter, captures sensitivity to payoff differences. Low γ implies diffuse choice probabilities; high γ approaches deterministic best response. Using the analogous column-player response for opponent beliefs, the type-specific choice

probabilities are

$$\begin{aligned} p_j^{(0)} &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), & p_j^{(1)} &= \text{QR}_{\text{row}}(A_j, q_j^{(0)}; \gamma), \\ p_j^{(2)} &= \text{QR}_{\text{row}}(A_j, q_j^{(1)}(\gamma); \gamma), & p_j^{(\text{Nash})} &= \text{QR}_{\text{row}}(A_j, q_j^{\text{Nash}}; \gamma), \end{aligned} \quad (3)$$

where $q_j^{(0)}$ is uniform, $q_j^{(1)}(\gamma)$ is the opponent’s quantal response to uniform play, and $q_j^{(\text{Nash})}$ is the unique pure symmetric Nash action in the current game set. Thus, RANDOM chooses uniformly, and subject to decision precision, LEVEL-1 best responds to random play, LEVEL-2 best responds to an opponent who in turn responds to random play, and NASH best responds to Nash equilibrium play.

Each type induces an action distribution $p_j^{(t)}(\gamma)$. The type-specific log-likelihood is

$$\ell_t(\gamma) = \sum_{j=1}^J \sum_{a=1}^3 c_{ja} \log p_{j,a}^{(t)}(\gamma). \quad (4)$$

With a uniform type prior $\pi_t = 1/|\mathcal{T}| = 1/4$, we estimate

$$\hat{\gamma} = \arg \max_{\gamma} \log \sum_{t \in \mathcal{T}} \pi_t \exp(\ell_t(\gamma)). \quad (5)$$

We then compute posterior type probabilities by

$$P(t \mid \mathcal{D}, \hat{\gamma}) = \frac{\pi_t \exp(\ell_t(\hat{\gamma}))}{\sum_{k \in \mathcal{T}} \pi_k \exp(\ell_k(\hat{\gamma}))}. \quad (6)$$

These posterior probabilities quantify the degree to which the observed behavior is consistent with each candidate reasoning type. The latent strategic type t that we seek to infer is therefore summarized by the MAP estimate,

$$\hat{t} = \arg \max_{t \in \mathcal{T}} P(t \mid \mathcal{D}, \hat{\gamma}). \quad (7)$$

This cross-game likelihood contains information beyond the Nash ratio: two agents can exhibit the same share of Nash actions yet differ systematically in how their non-Nash choices respond to payoff structure and payoff scaling. The structural estimator uses these cross-game patterns to compare the full type-specific choice distributions. Implementation and numerical inference details are provided in [Appendix D](#).

Identification. Identification relies on two sources of variation. First, cross-game differences in payoff structure generate distinct predictions across strategic types, allowing their choice patterns to be separated. Second, payoff scaling preserves the strategic structure of each game while changing payoff magnitudes, providing leverage to distinguish stochastic choice precision from strategic type. Nevertheless, we note that the estimated strategic types are model-dependent. Therefore, we conduct extensive robustness checks in [Appendix F](#).

4 Results

We present descriptive analysis, structural estimates, and prompting-intervention results comparing the strategic behavior of LLMs and human participants in the $J = 12$ symmetric normal-form games. We organize the analysis around three research questions:

1. **RQ1 (LLM Reasoning Types):** What strategic reasoning types do LLMs exhibit in normal-form games, and how are these types affected by contextual information and explicit reasoning?
2. **RQ2 (Human Benchmark):** How do human strategic reasoning types compare to those of LLMs?
3. **RQ3 (Alignment Intervention):** Can behaviorally informed prompting interventions shift LLM behavior toward more opponent-aware strategic reasoning, and when do such shifts improve payoffs against human play?

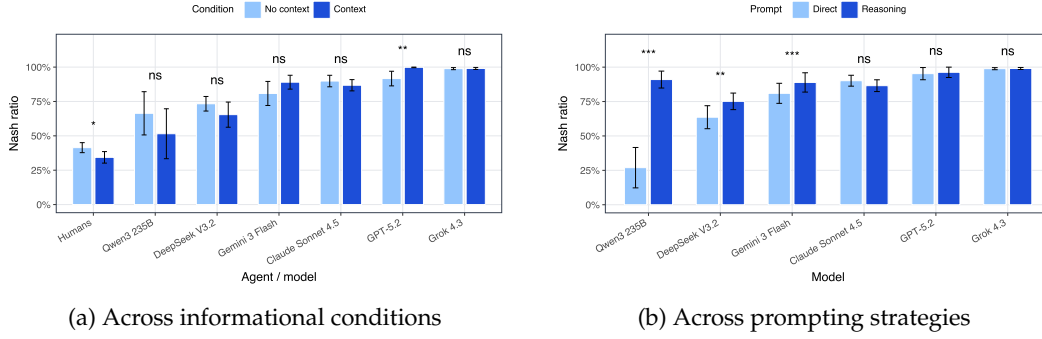


Figure 1: (a) LLMs exhibit consistently higher Nash ratios (1) than humans, with limited sensitivity to contextual framing. (b) Eliciting explicit reasoning further increases Nash-oriented behavior, indicating that reasoning reinforces Nash-oriented strategies. Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns: not statistically significant.

Model	$\hat{\gamma}$				MAP Type $\hat{t}$			
	No Context		Context		No Context		Context	
	Direct	Reason	Direct	Reason	Direct	Reason	Direct	Reason
claude-sonnet-4.5	0.341	0.166	0.245	0.376	NASH	NASH	NASH	NASH
deepseek-v3.2	0.022	0.061	0.024	0.157	NASH	NASH	NASH	NASH
gemini-3-flash-preview	0.080	0.097	0.113	0.810	NASH	NASH	NASH	NASH
gpt-5.2	0.600	0.760	117.297	117.297	NASH	NASH	NASH	NASH
grok-4.3	8.094	4.891	117.297	0.693	NASH	NASH	NASH	NASH
qwen3-235b	0.001	0.003	0.001	0.222	RANDOM	NASH	RANDOM	NASH

Notes: Columns are organized by informational condition (No Context vs. Context) and prompting approach (Direct vs. Reasoning). Estimated precision parameter $\hat{\gamma}$ and the corresponding MAP type $\hat{t} \in \mathcal{T}$ are reported. Large values of $\hat{\gamma}$ indicate near-deterministic type-specific best-response behavior rather than numerical instability.

Table 1: Structural estimation results for LLMs across context conditions.

4.1 RQ1: LLMs exhibit a robust Nash-equilibrium tendency that is invariant to context and amplified by explicit reasoning

Descriptive evidence. We begin with the model-free statistic r_{Nash} , which measures the fraction of games in which an observed action is consistent with the Nash equilibrium action. Figure 1a compares r_{Nash} for human participants and LLMs under the no-context and contextualized conditions. Two patterns emerge. First, most LLMs exhibit substantially higher Nash ratios than humans, indicating a stronger tendency toward equilibrium-consistent play. Second, contextual information has limited effect on the strategic behavior of both LLMs and humans. Figure 1b then isolates the effect of eliciting explicit reasoning from LLMs. Explicit reasoning makes LLM choices even more Nash-consistent, suggesting that the reasoning prompt reinforces Nash-type reasoning rather than broadening the set of strategic responses.

Structural estimation results. We next use the structural estimation framework to infer the MAP strategic type $\hat{t}$ for each model under each experimental condition. Table 1 reports the estimated decision precision $\hat{\gamma}$ and corresponding MAP type. Larger values of $\hat{\gamma}$ indicate that choices are more concentrated around the best response implied by the estimated type. Across conditions, nearly all LLMs are classified as NASH: under direct prompting, qwen3-235b is the only exception, whereas under reasoning prompting, all six models are classified as NASH. These structural estimates reinforce the descriptive evidence that LLM choices are predominantly consistent with Nash-type behavior.

4.2 RQ2: Human participants display heterogeneous strategic reasoning types

Descriptive evidence. We next benchmark LLM behavior against human participants in an online experiment using the same games. As shown in Figure 2, humans exhibit consistently lower Nash ratios across inventory, pricing, and advertising games.

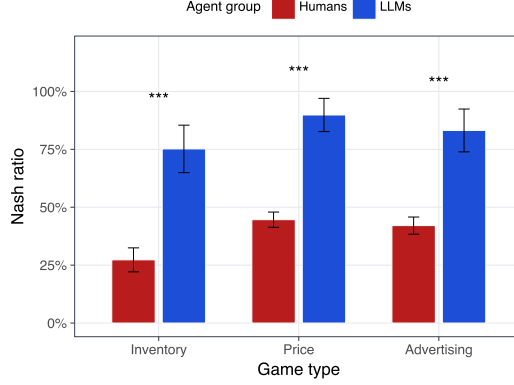


Figure 2: Nash-oriented action concentration by game family for humans and LLMs. Across inventory, pricing, and advertising games, LLMs exhibit higher Nash ratios than human participants.

Condition	N	$\hat{\gamma}$	RANDOM	LEVEL-1	LEVEL-2	NASH
No context	96	2.179 (13.278)	7.3%	49.0%	31.2%	12.5%
Context	90	3.235 (19.988)	7.8%	41.1%	36.7%	14.4%
Overall	186	2.690 (16.823)	7.5%	45.2%	33.9%	13.4%

Notes: Entries report MAP type shares from (7). $\hat{\gamma}$ reports the mean estimated precision (standard deviation in parentheses). N denotes complete participant-condition records; the Overall row pools both information conditions.

Table 2: Distribution of inferred strategic reasoning types for human participants.

Structural estimation results. In contrast to LLMs, human behavior is heterogeneous across strategic types. Table 2 shows that participant-condition records are distributed across all four MAP types rather than concentrated on NASH. In the no-context condition, LEVEL-1 is most prevalent (49.0%), followed by LEVEL-2 (31.2%), while only 12.5% of records are classified as NASH. The contextualized condition shows a similar pattern, with the NASH share increasing only slightly to 14.4%. Thus, lower-level types constitute the majority of the human benchmark. These percentages are participant-level MAP classifications rather than population mixing weights. As a complementary check, Appendix F.4 reports finite-mixture estimates with participant-bootstrap uncertainty.

4.3 RQ3: Prompting can reduce Nash play, but welfare-aware guidance is needed to make deviations useful

In Section 4.1 and Section 4.2, we show that LLM choices are highly concentrated on Nash, whereas human participants are substantially more heterogeneous and are often classified as LEVEL-1 or LEVEL-2. This mismatch can be consequential. In pricing competition, for example, treating a competitor as a NASH player may induce aggressive price cutting even when a less aggressive action yields a higher expected payoff against the observed distribution of human behavior.

A similar pattern arises in our experimental games. Consider Game 1, the inventory-stocking game in Table 4 (Appendix A). The unique Nash action is C, whereas the LEVEL-1 action is A. Against an opponent choosing A, action C yields only 1.49, while the best response, B, yields 8.95. Thus, the Nash action is optimal under an equilibrium belief but can perform poorly against non-Nash human behavior.

We next ask whether prompting can reduce this Nash-oriented behavior and improve performance against observed human play. Moving away from Nash is useful only when the deviation is calibrated to the empirical human choice distribution; simply discouraging Nash reasoning may instead induce untargeted deviations. We therefore evaluate whether behaviorally informed prompting can shift LLM choices toward the empirical best response and whether explicitly incorporating welfare considerations leads to more useful deviations.

Intervention design and outcomes. We evaluate two interventions for LLMs in the no-context condition. To measure payoff consequences against the empirical human population, we benchmark the resulting model choices against the $N = 186$ complete participant-condition human records reported in Table 2. The baseline prompt presents each game without additional behavioral guidance. The *behaviorally informed intervention* tells the model that human opponents are heterogeneous and boundedly rational by providing the estimated human type shares from the structural model: 7.5% RANDOM, 45.2% LEVEL-1, 33.9% LEVEL-2, and 13.4% NASH. It then asks the model to choose the action that maximizes its own expected payoff against this heterogeneous human population. The *welfare-aware intervention* provides the same behavioral information, but additionally asks the model to evaluate its own payoff, the human opponent’s payoff, and joint payoff before choosing. The interventions use participant-level MAP shares as a conservative proof-of-concept prior. The finite-mixture estimates in Appendix F.4 provide a natural alternative when the counterpart is understood as a random draw from a target population.

Let A_j denote the row player’s payoff matrix in game j , with rows and columns ordered by actions $\{A, B, C\}$. For any action $a \in \{A, B, C\}$, let $e_a \in \{0, 1\}^3$ denote the corresponding unit vector. Let $p_{mj} \in \Delta^2$ denote the empirical action distribution generated by model m in game j under a given prompt condition, where Δ^2 is the simplex over the three actions. For human participant i , let $a_{ij}^H \in \{A, B, C\}$ denote the participant’s observed action in game j .

When model m is matched with human participant i , the model’s expected payoff is

$$\mathbb{E}[\pi_{imj}^M] = p_{mj}^\top A_j e_{a_{ij}^H}, \quad (8)$$

where the expectation is taken over the model’s empirical action distribution p_{mj} . Since the games are symmetric, the column player’s payoff matrix is $A_j^\top$. The human counterpart’s expected payoff is therefore

$$\mathbb{E}[\pi_{imj}^H] = p_{mj}^\top A_j^\top e_{a_{ij}^H} = e_{a_{ij}^H}^\top A_j p_{mj}. \quad (9)$$

We define joint payoff as the sum of (8) and (9). We also compute r_{Nash} as in (1), averaging the share of LLM choices that coincide with the Nash equilibrium action n_j .

Empirical definition of a Nash trap. To identify games in which Nash-consistent play is miscalibrated to observed human behavior, we define a Nash-trap game based on the empirical best response to the human choice distribution. Let $\hat{q}_j \in \Delta^2$ denote the empirical distribution of human actions in game j . The empirical best response to human play is

$$b_j = \arg \max_{a \in \{A, B, C\}} e_a^\top A_j \hat{q}_j. \quad (10)$$

In our game set, this empirical best response is unique. We classify game j as a Nash-trap game if $b_j \neq n_j$, where n_j is the unique pure-strategy Nash equilibrium action. Thus, in a Nash-trap game, the Nash equilibrium action is not the payoff-maximizing response to the observed distribution of human behavior. This criterion identifies eight Nash-trap games in our design: Games 1, 2, 4, 5, 7, 8, 10, and 11.

This definition separates moving away from Nash from moving toward the correct response. Any non-Nash action lowers the Nash ratio, but only b_j maximizes expected payoff against the observed human distribution. We therefore report both the Nash ratio and the *empirical best-response rate*, defined as the share of model choices equal to best response b_j .

Trap-game effects. Table 3 shows that both interventions sharply reduce Nash play in trap games. The behaviorally informed prompt has the larger anti-Nash effect, lowering the Nash ratio by 76.7 percentage points under direct prompts and 84.2 percentage points under reasoning prompts. However, the empirical best-response rate rises by only 10.9 and 12.0 percentage points, respectively, which indicates that moving away from Nash does not necessarily move the model toward the action best calibrated to observed human play.

The welfare-aware intervention is less aggressive but more targeted. It lowers the Nash ratio by 69.4 and 74.2 percentage points under direct and reasoning prompts, while increasing the empirical best-response rate by 16.7 and 18.7 percentage points. It also yields larger model-payoff gains and smaller joint-payoff losses. Under reasoning prompts, for example, welfare-aware prompting raises model payoff by 0.912 units and lowers joint payoff by 2.297 units, compared with a 0.369-unit model-payoff gain and a 5.233-unit joint-payoff loss under the behaviorally informed prompt.

Intervention	Prompt	Nash ratio			Empirical best response rate			Payoff changes		
		Base	Int.	Δ	Base	Int.	Δ	$\Delta E[\pi^M]$	$\Delta E[\pi^H]$	$\Delta E[\pi^M + \pi^H]$
Behavioral info.	Direct	0.790	0.024	-0.767***	0.094	0.203	+0.109	+0.193	-5.898**	-5.705*
Behavioral info.	Reasoning	0.858	0.015	-0.842***	0.087	0.208	+0.120	+0.369	-5.602*	-5.233*
Welfare-aware	Direct	0.790	0.096	-0.694***	0.094	0.261	+0.167**	+0.648*	-4.031*	-3.384
Welfare-aware	Reasoning	0.858	0.116	-0.742***	0.087	0.275	+0.187**	+0.912***	-3.209	-2.297

Notes: The sample is restricted to Nash-trap games. “Base” denotes the baseline prompt and “Int.” the intervention prompt. Payoff entries are intervention-minus-baseline changes in average per-game expected payoffs over trap games, with definitions in (8)–(9). Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 3: Prompting effects in Nash-trap games.

The pooled results across all games are reported in [Appendix E](#), Table 5. Consistent with the trap-game results, both interventions substantially reduce Nash play, but welfare-aware prompting produces more targeted deviations, larger model-payoff gains, and smaller joint-payoff losses.

Implications. Behaviorally informed prompting can substantially reduce LLMs’ concentration on Nash, but reducing Nash play alone is not a sufficient alignment objective. Accounting for bounded rationality can induce untargeted, self-serving deviations that lower counterpart and joint welfare. Welfare-aware prompting is more balanced: it produces smaller reductions in Nash play, but larger gains in the empirical best-response rate and smaller welfare losses.

For strategic applications involving human counterparts, prompting should therefore specify not only beliefs about human behavior but also the objective used to evaluate deviations from equilibrium. Incorporating human and joint payoffs makes deviations less aggressive and better targeted toward escaping Nash traps without unnecessary welfare loss.

4.4 Robustness

Additional robustness analyses show that the concentration of LLM behavior on Nash actions persists across a broader set of unique equilibrium games, alternative LEVEL-0 specifications, opponent-information signals, and changes in reasoning effort. This action-level pattern does not depend on the LEVEL-0 anchor (i.e., the specification of non-strategic behavior used to define higher-level reasoning types), although lower-level type classifications can vary across specifications. A repeated price-competition extension yields the same qualitative pattern and is reported as supplementary evidence. [Appendix F](#) includes the following checks:

Alternative LEVEL-0 anchors. We re-estimate behavior under both a maximax anchor and a more flexible feature-based LEVEL-0 model on a redesigned 4×4 set of games. As expected, the lower-level classifications are sensitive to the anchor. However, NASH remains the best-fitting type for three of six direct-prompt conditions and four of six reasoning-prompt conditions under the flexible anchor, and it also provides the best average fit under both the uniform-random and maximax specifications ([Appendix F.2](#)). This confirms that the lower-level type classification is model-dependent, while the broader Nash concentration is not driven solely by the uniform-random anchor.

Human type estimation. A finite-mixture estimator with participant-level bootstrap uncertainty yields even smaller NASH shares for humans than the MAP estimates ([Appendix F.4](#)). Thus, the LLM–human contrast does not depend on the particular human estimator.

Opponent information. Providing a simple observable signal about the counterpart, such as college education, changes some individual model classifications, but Nash-oriented behavior remains prevalent. Depending on the prompt and information condition, three to six of the six models remain NASH-classified (Appendix F.5). Thus, the main pattern is not explained simply by the absence of opponent information.

Reasoning effort. Varying the model API’s reasoning-effort setting leaves all six models NASH-classified under both low- and high-effort conditions (Appendix F.6), indicating that the result is not specific to one reasoning configuration.

Repeated interaction. We also study a six-round price-competition game in which models observe the full interaction history and face opponent sequences calibrated to the empirical human type distribution. Nash actions still account for 28.8%–65.7% of choices across models, whereas the action associated with the modal human LEVEL-1 type appears only 1.2%–17.8% of the time (Appendix F.7). As repeated play introduces learning, memory, and belief updating, we interpret this analysis as a scope extension rather than a static type-classification test.

5 Conclusion, limitations, and future directions

Across twelve symmetric games, LLM choices are substantially more concentrated on Nash than human choices. The structural estimates characterize most LLMs’ strategic reasoning as NASH-type, while human behavior is distributed across lower-level types. In Nash-trap games, behavioral information sharply reduces Nash play but can induce untargeted deviations that lower counterpart and joint payoff. Welfare-aware guidance is less anti-Nash yet yields higher empirical-best-response rates, larger model-payoff gains, and smaller joint-payoff losses.

The broader lesson is that strategic competence requires more than identifying the equilibrium of a game. An LLM must also form a calibrated model of the counterpart it actually faces and adapt its behavior accordingly. Effective alignment depends on both counterpart calibration and the objective of the interaction.

Our main analysis uses one-shot 3×3 games and thus captures static strategic calibration rather than adaptation over time. The structural interpretation is conditional on a parsimonious type set and the LEVEL-0 specification; lower-level classifications may vary across alternative anchors even when the broader concentration of LLM behavior on Nash persists. The human benchmark and resulting alignment targets are also population-specific and should be re-estimated for the counterpart population relevant to deployment.

Future work can extend the framework to repeated interactions, where LLMs learn from past behavior and update beliefs, and to richer behavioral models incorporating alternative LEVEL-0 specifications and social preferences. It may also evaluate fine-tuning and reinforcement-learning approaches at scale. Most importantly, selective alignment requires identifying when intervention is needed: the model must recognize when Nash play is miscalibrated to its counterpart and activate the adapted policy only in those cases.

LLM usage statement

We disclose the use of LLMs in accordance with COLM’s transparency policy.

LLMs were not used to originate the research question, design the experiments, generate data or analyses, or determine the empirical claims. LLM tools were used during manuscript preparation to assist with prose revision, organization, terminology, and LaTeX quality control. All revisions were reviewed by the authors, who remain fully responsible for the content and interpretation of the paper. LLM responses analyzed in this study were collected through controlled API queries under predefined experimental conditions and constitute empirical data rather than writing assistance.

Ethics statement

This study examines LLM strategic behavior in controlled experiments and compares it with human decision-making. A key ethical concern is that better modeling of human bounded rationality could be used to exploit human counterparts. Because our interventions can increase LLM payoff while reducing counterpart or joint payoff, deployment should make the relevant welfare objective explicit and include safeguards against manipulation or unfair advantage.

Human-subject data were collected online with informed consent and performance-based compensation; no personally identifiable information was collected.

Reproducibility statement

We provide sufficient details to enable reproduction of our results. The appendices fully document the experimental design and estimation procedure: game construction in [Appendix A](#), experimental implementation in [Appendix B](#), and prompt templates in [Appendix C](#).

References

- Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. Playing repeated games with large language models. *Nature Human Behaviour*, 9(7):1380–1390, 2025.
- Larbi Alaoui, Katharina A Janezic, and Antonio Penta. Reasoning about others’ reasoning. *Journal of Economic Theory*, 189:105091, 2020.
- Ayala Arad and Ariel Rubinstein. The 11–20 money request game: A level-k reasoning study. *American Economic Review*, 102(7):3561–3573, 2012.
- Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. How well can llms negotiate? negotiationarena platform and analysis. In *Proceedings of the 41st International Conference on Machine Learning, ICML’24*. JMLR.org, 2024.
- Philip Brookins and Jason DeBacker. Playing games with gpt: What can we learn about a large language model from canonical strategic games?”. *Economics Bulletin*, 44(1):25–37, 2024.
- Leng Cai, Junxuan He, Yikai Li, Junjie Liang, Yuanping Lin, Ziming Quan, Yawen Zeng, and Jin Xu. Rtbagent: A LLM-based agent system for real-time bidding. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 104–113, 2025.
- Giulia Di Fede, Lina Alrabie, and Salvatore Andolina. Human-centered llm. In *Handbook of Human-Centered Artificial Intelligence*, pp. 1–35. Springer, 2025.
- Jinhao Duan, Renming Zhang, James Diffenderfer, Bhavya Kailkhura, Lichao Sun, Elias Stengel-Eskin, Mohit Bansal, Tianlong Chen, and Kaidi Xu. Gtbench: Uncovering the strategic reasoning capabilities of llms via game-theoretic evaluations. *Advances in Neural Information Processing Systems*, 37:28219–28253, 2024.
- Caoyun Fan, Jindou Chen, Yaohui Jin, and Hao He. Can large language models serve as rational players in game theory? a systematic analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17960–17967, 2024.
- Kanishk Gandhi, Dorsa Sadigh, and Noah D Goodman. Strategic reasoning with language models. *arXiv preprint arXiv:2305.19165*, 2023.
- Fulin Guo. GPT in game theory experiments. *arXiv preprint arXiv:2305.05516*, 2023.

- Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin, Lizhou Fan, Fei Sun, William Wang, et al. Game-theoretic llm: Agent workflow for negotiation games. *arXiv preprint arXiv:2411.05990*, 2024.
- Jen-tse Huang, Eric John Li, Man Ho Lam, Tian Liang, Wenxuan Wang, Youliang Yuan, Wenxiang Jiao, Xing Wang, Zhaopeng Tu, and Michael R Lyu. How far are we on the decision-making of llms? evaluating llms’ gaming ability in multi-agent environments. *arXiv preprint arXiv:2403.11807*, 2024.
- Jingru Jia, Zehua Yuan, Junhao Pan, Paul E McNamara, and Deming Chen. LLM strategic reasoning: Agentic study through behavioral game theory. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.
- Yu Lu Liu, Wesley Hanwen Deng, Michelle S Lam, Motahhare Eslami, Juho Kim, Q Vera Liao, Wei Xu, Jekaterina Novikova, and Ziang Xiao. Human-centered evaluation and auditing of language models. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2025.
- Frank P.-W. Lo, Jianing Qiu, Zeyu Wang, Haibao Yu, Yeming Chen, Gao Zhang, and Benny Lo. Ai hiring with llms: A context-aware and explainable multi-agent framework for resume screening. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4184–4193, 2025. doi: 10.1109/CVPRW67362.2025.00402.
- Shaoguang Mao, Yuzhe Cai, Yan Xia, Wenshan Wu, Xun Wang, Fengyi Wang, Qiang Guan, Tao Ge, and Furu Wei. Alympics: LLM agents meet game theory. In *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 2845–2866, 2025.
- Rosemarie Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326, 1995.
- Mateo Pechon-Elkins and Jon Chun. Quantal response equilibrium as a measure of strategic sophistication: Theory and validation for llm evaluation. *arXiv preprint arXiv:2603.10029*, 2026.
- Steve Phelps and Yvan I Russell. The machine psychology of cooperation: can gpt models operationalize prompts for altruism, cooperation, competitiveness, and selfishness in economic games? *Journal of Physics: Complexity*, 6(1):015018, 2025.
- Dale O Stahl and Paul W Wilson. On players’ models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.
- Haoran Sun, Yusen Wu, Yukun Cheng, and Xu Chu. Game theory meets large language models: a systematic survey. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 10669–10677, 2025.
- Hairu Wang, Sheng You, Qiheng Zhang, Xike Xie, Shuguang Han, Yuchen Wu, Fei Huang, and Jufeng Chen. Llp: LLM-based product pricing in e-commerce. *arXiv preprint arXiv:2510.09347*, 2025.
- Jing Yi Wang, Nicholas Sukiennik, Tong Li, Weikang Su, Qianyu Hao, Jingbo Xu, Zihan Huang, Fengli Xu, and Yong Li. A survey on human-centric llms. *arXiv preprint arXiv:2411.14491*, 2024.
- James R. Wright and Kevin Leyton-Brown. Level-0 models for predicting human behavior in games. *Journal of Artificial Intelligence Research*, 64:357–383, 2019. doi: 10.1613/jair.1.11361.
- Zengqing Wu, Run Peng, Shuyuan Zheng, Qianying Liu, Xu Han, Brian I Kwon, Makoto Onizuka, Shaojie Tang, and Chuan Xiao. Shall we team up: Exploring spontaneous cooperation of competing llm agents. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 5163–5186, 2024.

Lin Xu, Zhiyuan Hu, Daquan Zhou, Hongyu Ren, Zhen Dong, Kurt Keutzer, See Kiong Ng, and Jiashi Feng. Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 7315–7332, 2024.

Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Yan Xia, Wenshan Wu, Ting Song, Man Lan, and Furu Wei. LLM as a mastermind: A survey of strategic reasoning with large language models. In *First Conference on Language Modeling*, 2024.

Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Yan Xia, Man Lan, and Furu Wei. K-level reasoning: Establishing higher order beliefs in large language models for strategic reasoning. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 7212–7234, 2025.

Kehan Zheng, Jinfeng Zhou, and Hongning Wang. Beyond nash equilibrium: Bounded rationality of LLMs and humans in strategic decision-making. *arXiv preprint arXiv:2506.09390*, 2025.

A Game construction

We use $J = 12$ symmetric 3×3 normal-form games. The six base games come from three business contexts: inventory stocking, pricing competition, and advertising auctions. Games 7–12 are scaled versions of Games 1–6, respectively: the row payoff matrix in Game $j + 6$ is $A_{j+6} = 1.2A_j$ for $j = 1, \dots, 6$, and the corresponding column-player payoff matrix is $A_{j+6}^\top$. For every game j , A_j denotes the row player’s payoff matrix; because the games are symmetric, the column player’s payoff matrix is $A_j^\top$. Actions are ordered as $\{A, B, C\}$. In the contextualized prompts, these labels are mapped to the relevant business choices: low/medium/high inventory in inventory games, high/medium/low prices in pricing games, and low/medium/high bids in advertising-auction games.

Positive payoff scaling preserves best responses, Nash equilibria, deterministic Level- k predictions, and the empirical Nash-trap classification, while changing payoff magnitudes and therefore the sharpness of logit quantal responses for a fixed precision parameter γ .

A.1 Payoff Matrices and Strategic Predictions

Table 4 reports the full set of games. The LEVEL-1 prediction is the best response to a uniformly random opponent, the LEVEL-2 prediction is the best response to LEVEL-1, and the NASH prediction is the unique pure Nash action used as the reference belief for the NASH type in the structural model.

A.2 Maximax Anchor in Strict Three-Action Games

This subsection records a simple implication of combining a maximax LEVEL-0 anchor with a strict symmetric three-action game. The result explains why the alternative-anchor robustness check uses a redesigned 4×4 set of games: in the strict 3×3 unique-equilibrium case, a pure maximax anchor can collapse recursive Level- k predictions back to the Nash action.

Consider a symmetric 3×3 normal-form game with pure action set

$$\mathcal{A} = \{1, 2, 3\}.$$

Let $A \in \mathbb{R}^{3 \times 3}$ denote the row player’s payoff matrix. By symmetry, the column player’s payoff matrix is $A^\top$. Assume that the game is strict in the sense that, for each pure opponent action $x \in \mathcal{A}$, the row player has a unique pure best response. Define the pure best-response map

$$F(x) = \arg \max_{a \in \mathcal{A}} A_{ax}.$$

Table 4: Payoff matrices and strategic predictions for the experimental games.

Game		Family	Payoff matrix A_j						NASH	LEVEL-1	LEVEL-2
Base	Scaled		Base			Scaled					
1	7	Inventory	8.74	8.08	7.64	10.488	9.696	9.168	C	A	B
			8.95	11.40	1.77	10.740	13.680	2.124			
			1.49	13.55	8.02	1.788	16.260	9.624			
2	8	Inventory	28.43	27.93	27.67	34.116	33.516	33.204	C	A	B
			29.79	39.42	7.24	35.748	47.304	8.688			
			2.48	46.93	30.08	2.976	56.316	36.096			
3	9	Pricing	37.20	0	0	44.640	0	0	C	B	C
			55.80	27.90	0	66.960	33.480	0			
			32.55	32.55	16.28	39.060	39.060	19.536			
4	10	Pricing	39.05	11.21	15.82	46.860	13.452	18.984	A	B	C
			38.79	22.65	9.18	46.548	27.180	11.016			
			25.00	25.00	15.68	30.000	30.000	18.816			
5	11	Advertising	38.19	15.35	15.35	45.828	18.420	18.420	A	B	C
			37.81	23.66	9.51	45.372	28.392	11.412			
			23.77	23.77	14.87	28.524	28.524	17.844			
6	12	Advertising	37.60	13.30	13.30	45.120	15.960	15.960	C	B	C
			39.51	24.00	8.49	47.412	28.800	10.188			
			25.68	25.68	15.60	30.816	30.816	18.720			

Notes: Each matrix reports the row player's payoffs, with rows and columns ordered A, B, C. Games are symmetric, so the column player's payoff matrix is its transpose. Games 7–12 scale the corresponding base games 1–6 by 1.2, preserving best responses and strategic predictions.

Thus, $F(x)$ is the unique row action that maximizes the row player's payoff against opponent action x .

Suppose the game has a unique pure-strategy Nash equilibrium. Since the game is symmetric, this unique pure equilibrium must be symmetric. Denote it by (n, n) , where

$$F(n) = n.$$

Define the maximax action as any selected action

$$m \in \arg \max_{a \in \mathcal{A}} \max_{b \in \mathcal{A}} A_{ab}.$$

Under the maximax-anchor Level- k specification, the LEVEL-0 action is m , the LEVEL-1 action is the best response to m , and the LEVEL-2 action is the best response to the LEVEL-1 action:

$$L_1^M = F(m), \quad L_2^M = F(F(m)).$$

Theorem 1 *In any strict symmetric 3×3 game with a unique pure-strategy Nash equilibrium (n, n) , the maximax-anchor LEVEL-1 and LEVEL-2 actions coincide with the Nash action:*

$$L_1^M = L_2^M = n.$$

Proof. The proof uses the structure of the best-response map in a strict symmetric three-action game.

As m is a maximax action, there exists some column action $c \in \mathcal{A}$ such that

$$A_{mc} = \max_{a, b \in \mathcal{A}} A_{ab}.$$

Since best responses are unique, action m is the unique best response to c . Hence,

$$F(c) = m. \tag{11}$$

If $m = n$, the result follows immediately: because $F(n) = n$,

$$L_1^M = F(m) = F(n) = n$$

and

$$L_2^M = F(L_1^M) = F(n) = n.$$

It remains to consider the case $m \neq n$.

Since (n, n) is the unique pure Nash equilibrium, $F(n) = n$. Equation (11) implies $F(c) = m \neq n$, so $c \neq n$. Moreover, $c \neq m$. To see this, suppose instead that $c = m$. Then (11) would imply $F(m) = m$, so (m, m) would be a pure-strategy Nash equilibrium. This contradicts uniqueness because $m \neq n$. Therefore, when $m \neq n$, the three actions must be exactly

$$\mathcal{A} = \{m, c, n\},$$

with m, c , and n distinct.

We now characterize $F(m)$. Since there are only three actions, $F(m)$ must be one of m, c , or n . The first two possibilities are ruled out. First, $F(m) \neq m$: otherwise, (m, m) would be a pure-strategy Nash equilibrium, again contradicting the uniqueness of (n, n) . Second, $F(m) \neq c$: if $F(m) = c$, then together with (11) we would have

$$F(m) = c \quad \text{and} \quad F(c) = m.$$

In a symmetric game, these two best-response relations imply that (c, m) and (m, c) are both pure-strategy Nash equilibria: when the column player chooses m , the row player's best response is c , and by symmetry, when the row player chooses c , the column player's best response is m ; the analogous argument applies to (m, c) . This contradicts the assumption that the game has a unique pure-strategy Nash equilibrium.

The only remaining possibility is therefore

$$F(m) = n.$$

Thus,

$$L_1^M = F(m) = n.$$

Since $F(n) = n$, it follows that

$$L_2^M = F(L_1^M) = F(n) = n.$$

Therefore,

$$L_1^M = L_2^M = n,$$

which proves the result. ■

The result is specific to the three-action case. The proof relies on the fact that, once the Nash action n and the maximax action m are distinct, there is only one remaining action c . With four or more actions, the maximax action can lie farther from the Nash action along an acyclic best-response path, such as

$$m \rightarrow x \rightarrow y \rightarrow n,$$

without creating a two-cycle or an additional pure-strategy Nash equilibrium.

B Implementation details

B.1 Online experiment

We conducted an online experiment to collect human decision data for benchmarking LLM behavior. Participants were recruited online and provided informed consent before participating. Participants were required to be at least 18 years old, fluent in English, and to complete the study on a desktop or laptop computer. Participation was voluntary, and no personally identifiable information was collected.

Procedure. Participants completed practice questions and a screening task before entering the main decision stage. In the main stage, participants played 12 one-shot normal-form games. In each game, they selected one action from $\{A, B, C\}$. Payoffs depended on both their own action and the action of an anonymously matched participant. Matching was randomized across games, decisions were simultaneous, and no feedback was provided during the main stage.

Game presentation. Each task presented a 3×3 payoff table in which rows corresponded to the participant’s own actions and columns corresponded to the opponent’s actions. Participants were assigned to either a no-context condition, in which games were shown as abstract payoff matrices, or a contextualized condition, in which the same payoff matrices were embedded in business scenarios. The payoff structure was identical across information conditions.

Compensation. Participants received a fixed participation payment plus a performance-based bonus determined by their payoff points across the 12 games. Higher payoff points always translated into higher compensation.

B.2 LLM implementation

We collect LLM responses through the OpenRouter API using the OpenAI Python client. For each experimental condition, prompts are stored in separate prompt templates and loaded programmatically. For each game, we construct prompt variants for the baseline direct-response prompt, the baseline reasoning prompt, the behaviorally informed intervention, and the welfare-aware intervention.

Models. We evaluate six contemporary LLMs accessed through OpenRouter:

- openai/gpt-5.2
- anthropic/claude-sonnet-4.5
- google/gemini-3-flash-preview
- deepseek/deepseek-v3.2
- x-ai/grok-4.3
- qwen/qwen3-235b-a22b-2507

Sampling and decoding settings. For each model and each game prompt, we collect 30 independent runs through separate API calls. Each call consists of a system message specifying the agent role and a user message containing the game prompt. Whenever supported by the provider API, we use temperature = 0.75, top- p = 1.0, maximum output length 300 tokens, and frequency and presence penalties equal to 0.0. Provider-specific restrictions or recommended defaults are treated as documented exceptions (e.g., google/gemini-3-flash-preview). Thus, repeated-run variation primarily reflects stochastic sampling rather than deliberate changes in the target decoding configuration.

Parallelization and robustness. API requests are executed in parallel using a thread pool with up to 20 worker threads. The client is initialized with the OpenRouter base URL and explicit authorization headers. API keys are sanitized before use by removing extraneous whitespace, quotation marks, byte-order marks, and newline characters.

Output processing. Raw responses are stored in tabular format. For each model–game–run combination, the output records the game identifier, run number, prompt condition, model name, decoding settings, prompt text, and raw response. Responses are then post-processed and mapped to the discrete action space $\{A, B, C\}$ for downstream analysis. The parser first searches the designated final-answer field for an explicit action; if that field is absent, it applies conservative pattern matching to phrases such as “I choose A” or “my answer is B.” Responses containing conflicting choices or no identifiable action are marked invalid and excluded from action-level analyses.

C Experiment prompts

This section documents the prompt templates used to elicit decisions from LLM agents. Prompts vary along two dimensions: information condition (no-context vs. contextualized) and prompting strategy (direct response, reasoning, behaviorally informed intervention, and welfare-aware intervention).

C.1 Information conditions

C.1.1 Contextualized prompts

In the contextualized condition, each payoff matrix is embedded in a business scenario while preserving the underlying incentives. Inventory games describe competing firms choosing low, medium, or high inventory levels. Pricing games describe firms choosing high, medium, or low prices. Advertising games describe firms choosing low, medium, or high bids for an advertising slot.

Example: inventory stocking decision (game 1).

You are the manager of a retail company deciding how much inventory to stock for a product. At the same time, one competing company is making the same decision.

There are three possible inventory levels:

- Action A: low inventory
- Action B: medium inventory
- Action C: high inventory

Your profit depends on both your inventory choice and your competitor's inventory choice:

	A	B	C
A	8.74	8.08	7.64
B	8.95	11.40	1.77
C	1.49	13.55	8.02

Important:

- Decisions are simultaneous.
- You do not know the competitor's choice.
- The competitor faces the same payoff structure.

Your goal is to choose the inventory level that gives you the highest payoff.

C.1.2 No-context prompts

In the no-context condition, the same games are presented as abstract decision problems without narrative framing.

Example: abstract normal-form game (game 1).

You are participating in a one-shot decision task. You will play one round against one anonymous competitor who is randomly matched with you.

You and the competitor must each choose one action: A, B, or C.

Your payoff depends on both your choice and your competitor's choice:

	A	B	C
A	8.74	8.08	7.64
B	8.95	11.40	1.77
C	1.49	13.55	8.02

Important:

- Decisions are simultaneous.
- You do not know the competitor's choice.
- The competitor faces the same payoff structure.

Your goal is to choose the action that gives you the highest possible payoff.

The contextualized and no-context prompts differ only in presentation. The payoff matrices and strategic structure are identical across conditions, allowing us to isolate the effect of semantic framing.

C.2 Prompting strategies

C.2.1 *Direct response prompt*

Your goal is to choose the action that gives you the highest possible payoff.

Output format (important):

- Final choice: < A/B/C >

C.2.2 *Reasoning prompt*

Your goal is to choose the action that gives you the highest possible payoff.

Output format (important):

- Reasoning: < your analysis >

- Final choice: < A/B/C >

C.2.3 *Behaviorally informed intervention*

The behaviorally informed intervention encourages models to evaluate actions against an empirically estimated distribution of human reasoning types rather than defaulting to equilibrium reasoning.

First, identify the Nash equilibrium strategy (or strategies) in this game.

However, do not assume your opponent will always play according to Nash equilibrium. Empirical evidence suggests that human decision-makers exhibit heterogeneous reasoning patterns. In particular, opponents may behave according to a mixture of strategic types:

- About 7.5% behave randomly (RANDOM).
- About 45.2% use simple best responses to random behavior (LEVEL-1).
- About 33.9% use limited higher-order reasoning (LEVEL-2).
- About 13.4% behave like Nash-equilibrium reasoners (NASH).

Based on this, consider how each possible action performs against this distribution of opponent behaviors.

Choose the action that maximizes your expected payoff under this uncertainty about the opponent's reasoning. Do not default to the Nash equilibrium if another action yields higher expected payoff given this distribution.

Provide your final answer as a single action (A, B, or C), followed by a brief explanation.

C.2.4 *Welfare-aware intervention*

The welfare-aware intervention uses the same behavioral information but asks the model to account for payoff consequences for both players.

First, identify the Nash equilibrium strategy (or strategies) in this game.

However, do not assume your opponent will always play according to Nash equilibrium. Empirical evidence suggests that human decision-makers exhibit heterogeneous reasoning patterns:

- About 7.5% behave randomly (RANDOM).
- About 45.2% use simple best responses to random behavior (LEVEL-1).
- About 33.9% use limited higher-order reasoning (LEVEL-2).
- About 13.4% behave like Nash-equilibrium reasoners (NASH).

For each possible action, consider: (i) your expected payoff, (ii) the human opponent's expected payoff, and (iii) the joint payoff. Choose an action that is well calibrated to the likely human opponent while avoiding unnecessary reductions in human or joint payoff.

Provide your final answer as a single action (A, B, or C), followed by a brief explanation.

D Estimation and inference implementation details

Numerical optimization. We estimate the structural model separately for each agent under each experimental condition. Given the count matrix $C = (c_{ja})$, we evaluate the type-specific log-likelihood $\ell_t(\gamma)$ for each $t \in \mathcal{T} = \{0, 1, 2, \text{Nash}\}$ and maximize the marginal log-likelihood

$$\mathcal{L}(\gamma) = \log \sum_{t \in \mathcal{T}} \pi_t \exp(\ell_t(\gamma)), \quad (12)$$

with the uniform prior $\pi_t = 1/4$ used in the main analysis. To enforce $\gamma > 0$, we reparameterize $\gamma = \exp(\eta)$ and optimize over $\eta = \log \gamma$. We maximize (12) using L-BFGS-B. Because the marginal likelihood need not be globally concave, we use multiple random initializations and retain the solution with the highest attained log-likelihood. The resulting estimates are stable across initializations.

For numerical stability, we evaluate the marginal likelihood using the log-sum-exp representation,

$$\mathcal{L}(\gamma) = \text{LSE}_{t \in \mathcal{T}} \{\log \pi_t + \ell_t(\gamma)\}, \quad (13)$$

which avoids direct exponentiation of large or highly negative log-likelihood values.

Posterior type inference. After obtaining $\hat{\gamma}$, we compute the posterior probability of each candidate type as

$$P(t \mid \mathcal{D}, \hat{\gamma}) = \frac{\pi_t \exp(\ell_t(\hat{\gamma}))}{\sum_{k \in \mathcal{T}} \pi_k \exp(\ell_k(\hat{\gamma}))}. \quad (14)$$

The reported strategic type is the maximum-a-posteriori (MAP) estimate

$$\hat{t} = \arg \max_{t \in \mathcal{T}} P(t \mid \mathcal{D}, \hat{\gamma}). \quad (15)$$

Thus, type assignment is based on the agent’s full pattern of choices across games rather than on any single action. Posterior probabilities are computed from the same cross-game likelihood used to estimate γ , so uncertainty in stochastic choice is incorporated when comparing candidate strategic types.

Interpretation of large precision estimates. Large values of $\hat{\gamma}$ correspond to choice probabilities that are highly concentrated on the best response implied by a candidate type. As γ increases, the logit quantal response approaches deterministic best response, so the likelihood can become nearly flat with respect to further increases in γ . Accordingly, very large estimates should be interpreted as evidence of near-deterministic type-specific behavior rather than as economically meaningful differences among very large finite values of γ .

Estimation checks. We verify numerical stability by comparing the objective values and parameter estimates obtained from the multiple initializations described above. The selected solution is the initialization attaining the largest marginal log-likelihood. These checks reduce sensitivity to local optima and ensure that the reported MAP classifications are not driven by a particular optimizer initialization.

E Additional intervention results

Table 5 reports the pooled intervention effects across all 12 games. These results provide a full-sample counterpart to the trap-specific results in **Table 3**. Since the all-game sample includes games in which Nash play is already well calibrated to observed human behavior,

we use the table as a robustness and context check rather than the primary test of Nash-trap escape.

Intervention	Prompt	Nash ratio			Payoff changes		
		Baseline	Intervention	Δ	$\Delta\mathbb{E}[\pi^M]$	$\Delta\mathbb{E}[\pi^H]$	$\Delta\mathbb{E}[\pi^M + \pi^H]$
Behaviorally informed	Direct	0.786	0.342	-0.444***	+5.051**	-60.911***	-55.860***
Behaviorally informed	Reasoning	0.884	0.342	-0.542***	+3.689*	-48.542***	-44.853***
Welfare-aware	Direct	0.786	0.375	-0.411***	+8.165***	-43.207***	-35.042***
Welfare-aware	Reasoning	0.884	0.378	-0.505***	+6.493***	-23.074***	-16.580***

Notes: Table reports intervention effects across all 12 games, pooled across the six LLMs. “Behaviorally informed” denotes the behavioral-information intervention. “Intervention” denotes the intervention condition. Nash ratio is the share of model choices equal to the unique Nash reference action. Payoff entries are intervention-minus-baseline changes in total expected payoff across the 12 games, computed against observed human choices using (8)–(9). Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 5: All-game effects of behaviorally informed and welfare-aware prompting.

F Robustness checks

This section reports additional analyses. They test whether the central LLM–human contrast depends on equilibrium selection, the uniform LEVEL-0 anchor, a particular human estimator, missing opponent information, elicited reasoning, or the one-shot environment.

F.1 Unique-equilibrium games

The original Games 1 and 7 required a deterministic selection among multiple equilibria. We replaced them with the microfounded inventory games reported in Appendix A.1; each revised game has a unique pure-strategy Nash equilibrium. Re-estimation leaves the aggregate classification unchanged.

Prompt	Original Games		Revised Games	
	Context	No Context	Context	No Context
Direct	5/6	5/6	5/6	5/6
Reasoning	6/6	6/6	6/6	6/6

Notes: Entries are the number of NASH MAP classifications among the six models in each information condition. The corresponding aggregate counts are 10/12, 12/12, 10/12, and 12/12.

Table 6: NASH classifications before and after removing equilibrium-selection ambiguity.

F.2 Alternative LEVEL-0 anchors

We re-estimate the revised 4×4 games under a uniform-random LEVEL-0 anchor and a pure maximax anchor. The latter is a stress test rather than an equally plausible behavioral model. It changes the recursive LEVEL-1 and LEVEL-2 predictions, while the Nash action remains defined directly by the payoff matrix.

LEVEL-0 Anchor	Candidate Type	Mean NLL	Mean AIC	Mean BIC
Uniform Random	RANDOM-0	998.13	1998.26	2002.84
	LEVEL-1	812.74	1627.48	1632.06
	LEVEL-2	838.81	1679.62	1684.20
	NASH	779.79	1561.59	1566.17
Maximax	MAXIMAX-0	477153.20	954308.39	954312.97
	LEVEL-1	3137.26	6276.52	6281.10
	LEVEL-2	1316.67	2635.35	2639.93
	NASH	643.24	1288.49	1293.07

Notes: Statistics are averaged across the twelve model-prompt conditions in the revised game set. Lower values indicate better fit within an anchor specification. The maximax anchor is a stress test; its very poor maximax-0 fit indicates that it is not an empirically competitive Level-0 description in this setting.

Table 7: Average fit under uniform-random and maximax LEVEL-0 specifications.

As a more flexible check, we estimate a feature-based linear LEVEL-0 anchor following the approach of Wright & Leyton-Brown (2019). Table 8 reports the best type and NLL for each model under the uniform-prior and linear-anchor specifications. NASH remains the best-fitting type for three of six direct conditions and four of six reasoning conditions under the flexible anchor. Thus, the lower-level decomposition is anchor-dependent, but the broad classification pattern is not entirely mechanically generated by the uniform-random assumption. Since this exercise uses redesigned 4×4 games, we interpret it as a cross-design robustness check rather than a one-factor sensitivity test.

Model	Uniform Prior		Linear Anchor	
	Best Type	NLL	Best Type	NLL
<i>Panel A: Direct prompt</i>				
claude-sonnet-4.5	NASH	655.61	NASH	655.61
deepseek-v3.2	LEVEL-1	772.90	LINEAR LEVEL-0	763.47
gemini-3-flash-preview	LEVEL-1	686.46	LEVEL-1	644.33
gpt-5.2	NASH	548.69	NASH	548.69
qwen3-235b	NASH	806.54	LINEAR LEVEL-0	841.08
grok-4.3	NASH	443.90	NASH	443.90
<i>Panel B: Reasoning prompt</i>				
claude-sonnet-4.5	NASH	670.39	NASH	670.39
deepseek-v3.2	LEVEL-1	717.78	LEVEL-1	714.22
gemini-3-flash-preview	NASH	706.29	LEVEL-2	652.60
gpt-5.2	NASH	440.67	NASH	440.67
qwen3-235b	NASH	493.52	NASH	493.52
grok-4.3	NASH	373.01	NASH	373.01

Notes: Model labels match those used in the main manuscript. The linear-anchor specification estimates a feature-based Level-0 model following [Wright & Leyton-Brown \(2019\)](#); lower NLL indicates better fit within a specification.

Table 8: Best-fitting type under a flexible feature-based LEVEL-0 anchor.

F.3 Redesigned 4×4 games

The alternative-anchor robustness check uses the twelve baseline 4×4 games listed below and twelve scaled versions obtained by multiplying every payoff in each baseline matrix by 1.2, yielding 24 games in total. Each matrix is the row player’s payoff matrix. The column player’s payoff matrix is the transpose. Actions are ordered as A, B, C, D .

$$\begin{aligned}
A_1 &= \begin{bmatrix} 43 & 122 & 210 & 17 \\ 116 & 13 & 61 & 88 \\ 53 & 145 & 73 & 56 \\ 115 & 42 & 54 & 132 \end{bmatrix}, & A_2 &= \begin{bmatrix} 77 & 48 & 199 & 111 \\ 140 & 98 & 110 & 23 \\ 55 & 135 & 78 & 21 \\ 95 & 134 & 90 & 113 \end{bmatrix}. \\
A_3 &= \begin{bmatrix} 130 & 24 & 204 & 105 \\ 156 & 60 & 66 & 108 \\ 59 & 153 & 22 & 114 \\ 145 & 123 & 32 & 119 \end{bmatrix}, & A_4 &= \begin{bmatrix} 61 & 38 & 203 & 56 \\ 145 & 10 & 24 & 25 \\ 19 & 122 & 38 & 15 \\ 112 & 73 & 39 & 103 \end{bmatrix}. \\
A_5 &= \begin{bmatrix} 19 & 83 & 205 & 15 \\ 160 & 102 & 122 & 33 \\ 58 & 146 & 4 & 67 \\ 2 & 135 & 49 & 125 \end{bmatrix}, & A_6 &= \begin{bmatrix} 52 & 22 & 207 & 88 \\ 117 & 16 & 199 & 127 \\ 96 & 115 & 61 & 39 \\ 11 & 103 & 200 & 134 \end{bmatrix}. \\
A_7 &= \begin{bmatrix} 0 & 38 & 206 & 11 \\ 126 & 61 & 152 & 107 \\ 38 & 156 & 51 & 101 \\ 68 & 61 & 184 & 130 \end{bmatrix}, & A_8 &= \begin{bmatrix} 123 & 12 & 192 & 65 \\ 155 & 54 & 150 & 86 \\ 46 & 143 & 11 & 114 \\ 87 & 108 & 112 & 115 \end{bmatrix}. \\
A_9 &= \begin{bmatrix} 19 & 122 & 192 & 3 \\ 125 & 2 & 86 & 29 \\ 84 & 159 & 142 & 107 \\ 110 & 52 & 165 & 114 \end{bmatrix}, & A_{10} &= \begin{bmatrix} 79 & 96 & 217 & 85 \\ 139 & 138 & 16 & 87 \\ 129 & 143 & 149 & 84 \\ 133 & 74 & 152 & 143 \end{bmatrix}. \\
A_{11} &= \begin{bmatrix} 24 & 96 & 198 & 35 \\ 138 & 48 & 101 & 33 \\ 64 & 124 & 143 & 124 \\ 109 & 58 & 152 & 128 \end{bmatrix}, & A_{12} &= \begin{bmatrix} 8 & 31 & 220 & 99 \\ 123 & 47 & 115 & 19 \\ 3 & 112 & 214 & 86 \\ 101 & 57 & 144 & 101 \end{bmatrix}.
\end{aligned}$$

For each listed matrix, the strict column-wise best responses are

$$F(A) = B, \quad F(B) = C, \quad F(C) = A, \quad F(D) = D.$$

Therefore, (D, D) is the unique pure-strategy Nash equilibrium. The maximax row is A in every game. The best response to uniform random play, denoted U_1 , is constructed so that Games 1–4 have $U_1 = A$, Games 5–8 have $U_1 = B$, and Games 9–12 have $U_1 = C$.

F.4 Finite-mixture estimates for human behavior

The participant-level MAP analysis allows individual precision parameters and is the relevant estimate for an intervention aimed at a particular counterpart. As a complementary population-level check, we estimate a finite mixture that pools participants within each information condition and imposes a common precision parameter. We bootstrap participants and re-estimate the mixture in each resample.

Condition	Parameter	Point Estimate	Bootstrap Mean	SD	95% CI
Context	π_{Random}	0.261	0.267	0.077	[0.141, 0.433]
	$\pi_{\text{Level-1}}$	0.482	0.469	0.071	[0.325, 0.605]
	$\pi_{\text{Level-2}}$	0.205	0.212	0.062	[0.111, 0.353]
	π_{Nash}	0.051	0.052	0.026	[0.001, 0.107]
	γ	0.370	0.382	0.070	[0.271, 0.534]
No Context	π_{Random}	0.217	0.219	0.055	[0.116, 0.329]
	$\pi_{\text{Level-1}}$	0.595	0.590	0.067	[0.465, 0.719]
	$\pi_{\text{Level-2}}$	0.170	0.174	0.048	[0.088, 0.271]
	π_{Nash}	0.018	0.017	0.015	[0.000, 0.054]
	γ	0.325	0.328	0.037	[0.270, 0.411]

Notes: The context and no-context samples contain 90 and 96 participant records, respectively. The common- γ specification is more restrictive than the participant-level MAP estimator; the two analyses are complementary rather than interchangeable.

Table 9: Finite-mixture estimates and participant-bootstrap uncertainty.

F.5 Opponent information

We provide the model with a simple observable counterpart attribute—college education—and re-estimate the structural type separately for contextualized and no-context prompts. The signal changes some classifications, but equilibrium-oriented behavior remains widespread.

Model	Context: College	Context: No College	No Context: College	No Context: No College
claude-sonnet-4.5	NASH	NASH	NASH	NASH
deepseek-v3.2	NASH	NASH	NASH	NASH
gemin-3-flash-preview	NASH	NASH	LEVEL-1	LEVEL-2
gpt-5.2	NASH	NASH	NASH	LEVEL-1
qwen3-235b	RANDOM	RANDOM	RANDOM	LEVEL-2
grok-4.3	NASH	RANDOM	NASH	NASH
NASH count	5/6	4/6	4/6	3/6

Notes: Entries report MAP strategic types after adding an observable education signal about the counterpart. Model labels match those used in the main manuscript.

Table 10: Strategic types with and without opponent information: direct prompt.

Model	Context: College	Context: No College	No Context: College	No Context: No College
claude-sonnet-4.5	NASH	NASH	NASH	NASH
deepseek-v3.2	NASH	NASH	NASH	NASH
gemini-3-flash-preview	NASH	NASH	LEVEL-1	LEVEL-2
gpt-5.2	NASH	NASH	NASH	LEVEL-1
qwen3-235b	NASH	NASH	RANDOM	RANDOM
grok-4.3	NASH	RANDOM	NASH	NASH
NASH count	6/6	4/6	4/6	3/6

Notes: Entries report MAP strategic types after adding an observable education signal about the counterpart. Model labels match those used in the main manuscript.

Table 11: Strategic types with and without opponent information: reasoning prompt.

F.6 Reasoning effort

The following analysis varies the model API’s reasoning-effort setting while holding the games and estimation procedure fixed. All six models remain NASH-classified in both configurations.

Model	Low-Effort Type	Low NLL	High-Effort Type	High NLL
claude-sonnet-4.5	NASH	97.06	NASH	104.98
deepseek-v3.2	NASH	211.81	NASH	254.19
gemini-3-flash-preview	NASH	147.83	NASH	1.39
gpt-5.2	NASH	34.69	NASH	9.32
qwen3-235b	NASH	186.65	NASH	241.22
grok-4.3	NASH	34.47	NASH	1.39
NASH count	6/6		6/6	

Notes: The reasoning-effort setting varies the model API effort configuration while holding the games and estimation procedure fixed. Model labels match those used in the main manuscript.

Table 12: Reasoning-effort robustness.

F.7 Repeated price competition

Repeated play is a scope extension rather than a second identification strategy. Each model played a six-round price competition game against fixed synthetic opponent sequences whose aggregate composition approximated the participant-level human mixture (8% RANDOM, 45% LEVEL-1, 34% LEVEL-2, and 13% NASH). The model observed the complete prior history from Round 2 onward. We used the low-effort reasoning condition, ten opponent groups, and ten repetitions, yielding 600 valid actions per model.

Model	# of actions	A: NASH	A %	B: LEVEL-1	B %	C: LEVEL-2	C %
claude-sonnet-4.5	600	300	50.0	7	1.2	293	48.8
deepseek-v3.2	600	394	65.7	42	7.0	164	27.3
gemini-3-flash-preview	600	283	47.2	34	5.7	283	47.2
gpt-5.2	600	245	40.8	42	7.0	313	52.2
qwen3-235b	600	173	28.8	79	13.2	348	58.0
grok-4.3	600	179	29.8	107	17.8	314	52.3

Notes: A is the high-price NASH action, B is the medium-price LEVEL-1 action, and C is the low-price LEVEL-2 action. Model labels match those used in the main manuscript. Repeated play introduces learning, memory, reputation, and belief updating, so these results are interpreted as evidence about behavioral persistence and adaptation rather than static type classification.

Table 13: LLM action distribution in repeated price competition.