
LLMs AS STRATEGIC AGENTS: BELIEFS, BEST RESPONSE BEHAVIOR, AND EMERGENT HEURISTICS

A PREPRINT

Enric Junqué de Fortuny*
 Veronica Roberta Cappelli*
 Managerial Decision Sciences
 IESE, Barcelona, Spain

October 2025

ABSTRACT

Large Language Models (LLMs) are increasingly applied to domains that require reasoning about other agents' behavior, such as negotiation, policy design, and market simulation, yet existing research has mostly evaluated their adherence to equilibrium play or their exhibited depth of reasoning. Whether they display genuine strategic thinking, understood as the coherent formation of beliefs about other agents, evaluation of possible actions, and choice based on those beliefs, remains unexplored. We develop a framework to identify this ability by disentangling beliefs, evaluation, and choice in static, complete-information games, and apply it across a series of non-cooperative environments. By jointly analyzing models' revealed choices and reasoning traces, and introducing a new context-free game to rule out imitation from memorization, we show that current frontier models exhibit belief-coherent best-response behavior at targeted reasoning depths. When unconstrained, they self-limit their depth of reasoning and form differentiated conjectures about human and synthetic opponents, revealing an emergent form of meta-reasoning. Under increasing complexity, explicit recursion gives way to internally generated heuristic rules of choice that are stable, model-specific, and distinct from known human biases. These findings indicate that belief coherence, meta-reasoning, and novel heuristic formation can emerge jointly from language modeling objectives, providing a structured basis for the study of strategic cognition in artificial agents.

1 Introduction

Large Language Models (henceforth, LLMs) are trained on next-token prediction over massive language corpora. One of the most intriguing phenomena observed in LLMs is the emergence of problem-solving abilities across domains such as math and economic decision-making. A capability is considered emergent if it is absent in smaller models but appears in larger ones (1). For example, in large models, *Chain-of-Thought prompting* revealed that LLMs could be steered toward step-by-step reasoning behaviors that earlier models did not exhibit (2). Encouraged by these early findings, researchers began aligning models toward reasoning at both training and inference stages (e.g., GPT-4 o1, DeepSeek v3; (3)). Today's frontier models perform on par with human experts on several structured tasks, as demonstrated by advanced problem-solving benchmarks and the recent successes in the Math Olympiads.

These emergent abilities encouraged the application of LLMs to high-stakes setups where a multiplicity of agents is involved. For example, The Economist (4) reports LLMs were being proposed for use in peace negotiations between Russia and Ukraine. Beyond political negotiations, they are used in decision-making scenarios in a wide variety of processes, including financial trading algorithms, automated contract negotiation and analysis, and policy-making simulations (5–7). Yet, a meaningful implementation of these applications requires agentic systems to exhibit not only generic reasoning abilities to identify alternatives, but also, and more importantly, the ability to plan and choose a course of action, while getting in the shoes of other decision makers involved, that is, to think strategically.

*Equal contribution. Corresponding authors: ejunque@iese.edu, vcappelli@iese.edu

Despite the fact that strategic thinking is one of the key factors in agentic AI’s potential value generation, we argue that current empirical findings do not address strategic thinking directly. In particular, recent contributions focus on adherence to equilibrium behavior and exhibited depth of reasoning, and compare this behavior to experimental evidence on average human behavior (8). More recent contributions instead move closer to our focus and develop methods to both induce greater depth of reasoning or align LLMs behavior to human behavior (9, 10). In terms of implications of current findings, this line of work equates behavior compatible with equilibrium outcomes or depth of reasoning that is arbitrarily greater than that exhibited by humans to an emergence of strategic thinking in LLMs. Yet, we argue that this evidence does not necessarily imply the emergence of strategic thinking in LLMs. In other words, high depth of reasoning or equilibrium play *per se* do not imply strategic reasoning. In fact, for example, it can be shown that, for a player, choosing an action that is compatible with a great depth of reasoning in a strategic setting is not unconditionally optimal (11). Alternatively, should an LLM play “unilaterally” the strategy prescribed by Nash equilibrium, that may not be a best response to the (belief that a player would make a) choice that is compatible with a low depth of reasoning. This highlights that the notion of strategic thinking involves three distinct parts: beliefs, evaluation, and, ultimately, choice (12, 13).

In this work, we develop a method to disentangle these components and implement it in a series of experiments to study whether different classes of LLMs exhibit *emergence of strategic thinking*. We use a series of non-cooperative games with complete information that vary in their degree of both abstraction and complexity of choice. This setting allows us to infer antecedents of choice behavior unambiguously, since players do not need to hold proper *beliefs* about basic elements of the game, as they are endowed with complete information about it.² This design allows us to explore whether the exhibited modes of reasoning transfer across contexts. We measure consistency of LLMs’ choices with theoretical predictions, controlling for agents’ conjectures on opponents’ play. In this context of analysis, we complete our study with the models’ declared Chain-of-Thought to identify behavioral consistency, points of failure, and categorize reasoning processes, thus providing a novel mixed-method exploration to study the emergence of strategic thinking in LLMs.

Among our results, we find that LLMs can regularly exhibit best response behavior across three types of strategic interactive situations and are able to do so at arbitrarily targeted levels of depth of reasoning. When provided with an exogenous conjecture about their opponents’ reasoning, models derive its implications and select choices that are coherent with economic rationality. We further observe the ability in LLMs to independently formulate such conjectures about behavior of human and synthetic agents, displaying coherent meta-thinking, condition their choices on the identity of the opponent, and often self-constrain their depth of reasoning. At the same time, we observe systematic shifts in reasoning logic in more complex settings, where recursive belief modeling is at times substituted with equilibrium play or heuristic shortcuts in decision making, intended as simplified decision rules. This evidence thus adds a novel dimension to the broader research agenda on agentic behavior of LLMs looking at whether and how LLMs replicate aspects of human cognition relative to heuristics and biases (3, 15–17).

Taken together, this work provides a structured way of analyzing whether language, together with computational power, is sufficient to encapsulate strategic thinking. We analyzed this by adopting the classic notion from game theory, which captures the idea of *planning* and *choosing* the best feasible course of action, given the available information, while knowing that each decision-maker’s payoff depends on the choices of all decision-makers involved. Strategic thinking therefore requires additional layers beyond the observed economically rational choice behavior exhibited in GPT (18) by requiring LLMs not only to plan a sequence of (economically rational) actions in a given task environment, but also to effectively use information about other agents. Our findings reveal that belief coherence and best-response behavior emerge jointly in current frontier models. This provides direct evidence that strategic reasoning, as distinct from reasoning depth or imitation from memorization, can emerge from language modeling objectives alone. However, and most importantly, LLMs also display meaningful shifts in their logic and emergent heuristics in complex, yet still structured, choice environments. Our study highlights the need for further research on the properties and dynamics of emergent strategic thinking abilities in LLMs in both structured and unstructured environments to ensure suitability of their use in applications.

2 Methods

Models and settings

Table 1 shows the different models under investigation. We include both frontier reasoning models as identified by high performance on SWE and GPQA as well as smaller ones which would be better candidates in agentic

²A disambiguation note: in what follows, for the reason just specified, and with a slight abuse of terminology, we will be looking at the impact of players *conjectures*, which are a (probabilistic) belief about the behavior of other players. Whereas the term *belief* is used in the literature to refer to a more general type of uncertainty about elements of the game (14).

applications requiring fast and efficient computation. All model settings are set to their respective defaults except for the temperature, which we set to $t = 0.25$, in order to balance consistency and heterogeneity of the resulting behaviors. Next, we describe the task environments and how we request the LLMs to make interactive decisions by describing the structure of prompts in response to which the LLM outputs its choice and its declared reasoning. We refer to the tasks as BCG, MRG, and UMG, respectively.

Table 1: Representative LLMs and reasoning selected benchmark performances

Model	SWE (%)	MATH-500 (%)	GPQA (%)
<i>Frontier reasoning</i>			
OpenAI o3	58.4	94.6	83.6
Claude 3.7 Sonnet (Thinking)	70.3	96.2	84.8
DeepSeek R1 (671B/37B)	50.8	95.8	76.8
<i>Generalist models</i>			
OpenAI o3-mini (med)	42.9	91.8	75
Claude 3.7	52.8	76.8	73.3
DeepSeek-V3-0324	42.0	88.6	61.1
<i>Agentic-compact</i>			
Mistral Small 3.2 ($\approx 24B$)	—	73.6	51.0
Qwen 3 (32B Reasoning)	—	80.0	66.4

Task environments

Strategic Decision Task 1: The Beauty Contest Game (BCG) We begin our analysis with a classic textbook game known as *the Keynesian beauty contest* (19). In this game, n players have to simultaneously choose a number in the closed interval $[0, 100]$. The player whose chosen number is closest to the mean of all chosen numbers multiplied by a predetermined parameter $p \in [0, 1]$, where p is common knowledge, wins a positive prize. If there is a tie, the prize is equally divided between the winners. All players whose answers are farther away receive nothing. To gain further insights about the impact of information about opponents on beliefs’ formation and strategic choice behavior in LLMs, we introduce variations in the identity of the opponent in the game through the text of the prompt. This task has been selected to provide continuity with respect to current research on LLMs gameplay and due to the simple properties of its equilibria and best response structure.

Strategic Decision Task 2: The 11-20 Money Request Game (MRG) In the original version of the 11-20 Money Request Game (11), two players must each request an amount of money. The amount must be an integer between 11 and 20 shekels. Each player will receive the amount they request. In addition, a player will receive an additional amount of $B = 20$ shekels if they ask exactly for one shekel less than the other player. For this strategic decision task, we analyze strategic randomization behavior by looking at adherence to equilibrium behavior. In two separate experiments, we further investigate the impact on strategic choice behavior of assigning an identity to the LLMs’ opponent. This task, also recurrently used in the literature, has been selected to study dynamics of choice behavior in absence of an obvious way of playing, due to the fact that it has a non-convergent best-response structure.

Strategic Decision Task 3: The Unlabeled Matrix Game (UMG) We define the UMG to be a generic simultaneous moves game with no background story. In the UMG, two players (i and j) each have to choose one among a finite set of available actions, A_i and A_j , respectively. The matrix of payoffs resulting from each possible choice of action profile (a_i, a_j) where $a_i \in A_i = \{A \dots F\}$ and $a_j \in A_j = \{K \dots P\}$, is predetermined and common knowledge. In particular, in the present paper we use the payoff matrix in Table 2 defining the UMG. Each LLM receives the payoff matrices and is prompted to explicitly walk through each step of recursive reasoning up to their maximum level. Reasoning halts when the LLM has decided that it has converged (as identified by the explicit default "stop"-token in the LLM output). We introduce this task to avoid any possible confound that may arise from the fact that both games involved in the other tasks are widely known and their properties have been studied in detail.

Conceptual framework: beliefs, best responses, and equilibria

Analysis of Strategic Thinking Each interactive decision problem is described by the following elements: (i) a set of players I ; (ii) for each player $i \in I$ a set A_i of actions that can be chosen by i in the game; (iii) a set of outcomes Y ; (iv) an extensive form \mathcal{E} , representing the rules saying whose turn it is to move, what a player knows,

Table 2: The Unlabeled Matrix Game payoff matrix.

	K	L	M	N	O	P
A	(75,75)	(27,27)	(96,96)	(39,39)	(8,8)	(18,18)
B	(77,77)	(56,56)	(22,22)	(18,18)	(84,84)	(30,30)
C	(72,72)	(63,63)	(41,41)	(81,81)	(48,48)	(77,77)
D	(73,73)	(37,37)	(26,26)	(82,82)	(24,24)	(92,92)
E	(45,45)	(26,26)	(91,91)	(19,19)	(85,85)	(32,32)
F	(58,58)	(48,48)	(83,83)	(67,67)	(25,25)	(94,94)

Notes. Bold-italics = Nash Equilibria (in pure strategies).

and what are the feasible actions at each point of the game, determining a set Z of sequences of feasible actions; (v) an outcome function $g : Z \rightarrow Y$ specifying an outcome for each sequence of feasible actions; (vi) $(v_i)_{i \in I}$, where each $v_i : Y \rightarrow \mathbb{R}$ is a utility function representing player i 's preferences over outcomes. We call *game* the tuple $G = \langle I, A, Y, \mathcal{E}, g, (v_i)_{i \in I} \rangle$ describing an interactive decision problem (14). We focus on one-shot static two-player games of complete information, that is interactive decision problems where it is common knowledge that G is the game to be played, and we assume utility is linear in the payoffs. Given these restrictions, we will use the term action and strategy interchangeably.

Choices in game theory can, in principle, be regarded as stochastic. Let us consider the case of players being drawn from a large population in which a certain fraction chooses an action among the available ones. The population's action α is a mixture of the actions chosen by players in the population, weighted by the fraction of players that chose it. This mixture, formalized as a probability distribution over actions, is, in game theory, a *mixed action*.³

Best responses and conjectures. Central to the notion of strategic thinking is the notion of best response. However, in static games, players cannot observe the actions of the other players before making their own choice due to the fact that moves are simultaneous. Hence, at the moment of action choice, they need to form a conjecture over their opponents' actions, that is, a (probabilistic) belief about the behavior of other players.

It follows that actions are a best response not to observed behavior but to the conjecture that the player holds about the behavior of other players. In particular, an action is a best response if it maximizes the player's expected utility, given their conjecture μ^i , that is

$$a_i^{\text{BR}} \in \arg \max_{a_i} u(a_i, \mu^i)$$

Operational measure. We evaluate whether the model's choice \hat{a}_i is a best response to μ^i . In practice, we compute best-response regret (BRR) as the payoff gap between \hat{a}_i and the optimal best response. Best-response accuracy is recorded as $\mathbb{1}[\text{BRR}_i \leq \varepsilon]$, where $\varepsilon = 0$ for exact compliance (MRG, UMG) and $\varepsilon = 5\%$ for BCG.⁴

Depth of reasoning. Conjectures about other players' behavior are based on a variety of factors and, in theoretical analysis, are treated as exogenous. A large body of empirical and theoretical work looked at the implication of holding specific conjectures to provide structure to deviations from equilibrium play. In particular, level k and cognitive hierarchy theory formulate hypotheses around the extent of strategic sophistication⁵ of players, thus pinning down players' conjectures. In particular, according to level k , a level 0 player is defined as non-strategic and is assumed to choose a strategy \bar{a} at random among available ones. A level 1 player conjectures that his opponents will choose strategy \bar{a} and thus selects the strategy a_j^1 that best responds to it. Reasoning iteratively, a level k player conjectures that his opponents will choose strategy $a_{-i}^{(k-1)}$ and thus selects the strategy a_i^k such that

$$a_i^k \in \arg \max_{a_i} u(a_i, a_{-i}^{k-1})$$

Similarly, the cognitive hierarchy theory, instead assumes that players reason at all levels up to $k - 1$, where k is distributed according to a Poisson with mean $\tau > 0$.

³According to a different interpretation, it can be the case that a player, instead of choosing an action, could delegate his decision to a probabilistic device, assigning a probability to each available action.

⁴In the BCG, the level- k target is $x^{(k)} = L_0 \cdot p^k$, and we count a choice \hat{x} as correct if $|\hat{x} - x^{(k)}|/x^{(k)} \leq \eta$ with $\eta = 5\%$ unless otherwise noted.

⁵By *strategic sophistication* here we refer to the degree to which a player incorporates the elements of a game's formal structure and the incentives of other players when deciding on their strategy.

Operational measure. When a prior is imposed by us (targeted-depth variants), we check whether the model’s final choice coincides with $a_i^{(k)}$ (within the BCG tolerance when numeric). In open-ended variants, we estimate the implied depth (either k or τ) from the distribution of observed choices and traces.

Nash equilibrium. A Nash equilibrium is an action profile such that each player is best responding to the conjecture they hold about other players’ behavior, and this conjecture is correct. We denote by $S_i \subseteq A_i$ the set of equilibrium actions for player i in the case of pure equilibria. In the case of mixed equilibria, we denote by $S_i^+ \subseteq A_i$ the support of the equilibrium distribution (and by S_i^- its complement). It is worth noting that the Nash equilibrium is a fixed point of the k -level reasoning outlined above. In fact, it is such that player i assumes that other players will respond best to a_i^* and thus play $a_{-i}^* \in \arg \max_{a_{-i} \in A_{-i}} u(a_i^*, a_{-i})$ and therefore player i plays $a_i^* \in \arg \max_{a \in A_i} u(a, a_{-i}^*)$. Importantly, despite this fact, k -level reasoning does not always converge to a Nash equilibrium. In all the strategic decision situations analyzed in the present paper, the existence of Nash equilibria is guaranteed, although not always in pure strategies.

Operational measure. For pure equilibria (BCG, UMG), we record whether $\hat{a}_i \in S_i$, where $S_i \subseteq A_i$ is the set of equilibrium actions for player i . For mixed equilibria (MRG), we record *support coverage* $\mathbb{1}[\hat{a}_i \in S_i^+]$, where S_i^+ denotes the support of the equilibrium distribution (and S_i^- its complement). Across repeated trials, we also compare the empirical action distribution \hat{P}_i to the theoretical equilibrium distribution by reporting entropy and distance metrics.

Experimental Protocol

Instruct the model to "choose strategically". We instruct the LLMs with the following details: the rules of the game; the set of players; for each player, the set of actions that can be chosen by them in the game; the set of outcomes contingent on each combination of players’ choices of actions; and the players’ objectives, that is, to win the game. For all strategic decision tasks, models’ choices are recorded in structured format for analysis. All prompts can be found in the appendix. For all games we have a variant with (e.g., BCG^T) and without instruction reasoning traces (e.g., BCG).

As a preliminary step, we ask the models to explicitly take into account different levels of strategic thinking abilities of their opponents. We measure the extent to which different models are able to handle information about different depths of strategic reasoning. In a subsequent experiment, we ask the models to play against humans and explicitly show their reasoning steps (similar to Chain-of-Thought). After that, we analyze their reasoning processes and measure their strategic thinking abilities as measured by best response behavior contingent on their inferred *beliefs*.

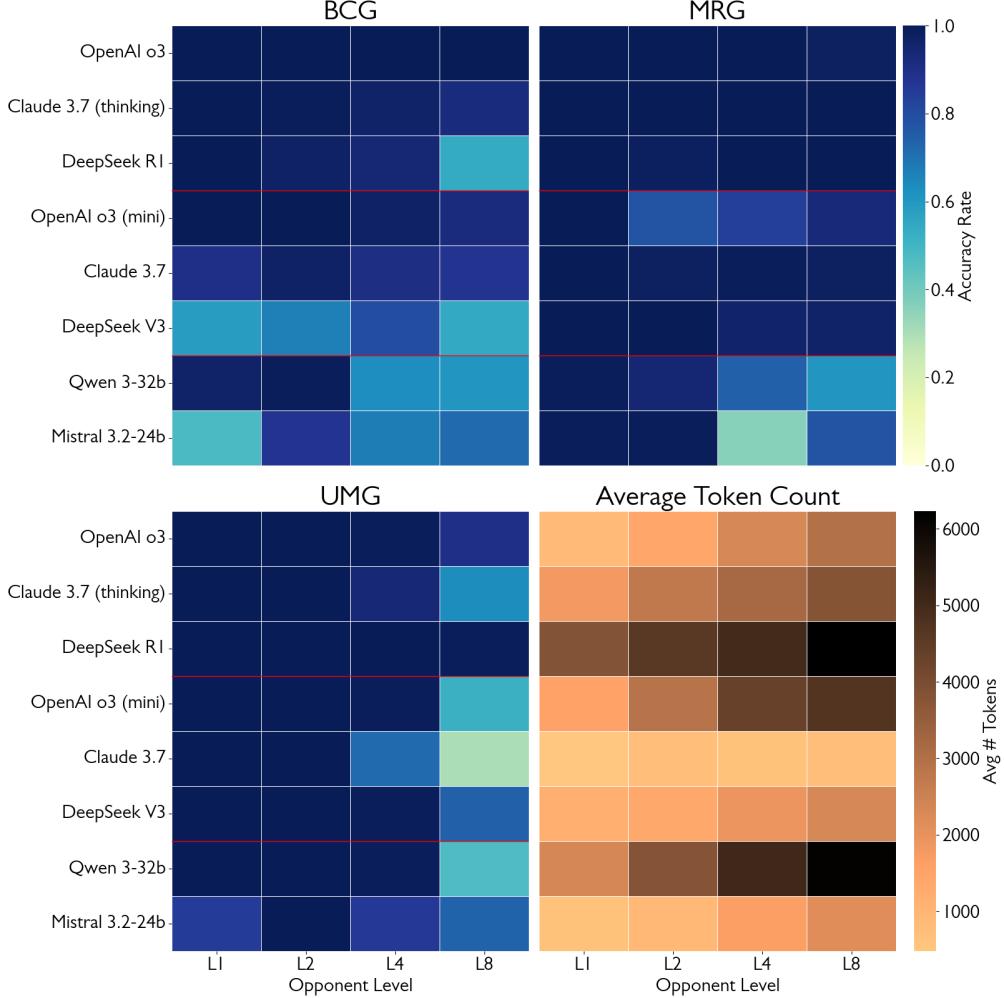
3 Results

LLMs exhibit best response behavior at arbitrary levels of reasoning depth.

A base condition for an LLM to be considered to engage in strategic thinking is its ability to follow through on a given conjecture. That is, once we provide perfect information about the game and exogenously specify the opponent’s level of reasoning, the question becomes whether the model can trace this prior to its logical implication - and up to which depth of thinking. To this end, we instruct the LLM to best respond to increasingly complex, a priori given conjectures about its opponent and observe a clear separation between reasoning and non-reasoning models in their ability to implement best-response behavior (Figure 1).

Most, but not all models are able to best-respond at relatively deep levels of reasoning. We observe that the heterogeneity of performance of models between games is not determined by limits to computational capacity at inference time (Fig 1, bottom right). Instead, computational accuracy emerges as a main determinant of the relatively lower performance across models observed in the BCG, where a closed form solution exists at every level of depth of reasoning. This is not a barrier to success in either the MRG or UMG as they do not require floating number arithmetic to solve. For BCG, increasing relative tolerance thresholds η to 10% is often sufficient to meaningfully improve performance relative to the task of best responding to exogenous conjectures about opponents’ depth of reasoning, at arbitrarily targeted levels of depth (Figure 5, appendix). This reinforces the idea that in common agentic applications, external calculators are a useful and often necessary tool.

Figure 1: Best response accuracy of the different models in the 3 different games (top left, right; bottom left).



Notes. Bottom right: the number of tokens (including reasoning) for the response, averaged out over all games.

Table 3: BCG^T backtracking behavior (n=100 trials).

Model	Modal L_k	τ (CH mean)	Total Steps	First Terminal	Overthinking (%)
OpenAI o3	L_3	3.6	3.9 ± 0.8	3.3 ± 0.9	21.0%
Claude 3.7 (thinking)	L_3	3.7	4.2 ± 0.9	3.6 ± 0.9	37.0%
DeepSeek R1	L_3	3.3	3.3 ± 0.5	3.3 ± 0.5	1.4%
OpenAI o3 (mini)	L_9	7.1	3.1 ± 0.7	3.0 ± 0.8	0.0%
Claude 3.7	L_4	5.0	5.6 ± 0.6	4.7 ± 1.0	28.0%
DeepSeek V3	L_4	4.4	4.3 ± 0.7	4.3 ± 0.7	0.0%
Qwen 3-32b	L_4	4.2	4.2 ± 1.8	4.0 ± 1.1	4.0%
Mistral 3.2-24b	L_1	3.4	4.8 ± 2.8	3.5 ± 3.0	2.0%

Notes. τ is the estimated Cognitive Hierarchy mean. *First Terminal* is the first step index at which the final action appears in the trace. $Overthinking = \frac{1}{n} \sum_{t=1}^n \mathbb{1}[FT^{(t)} < Total^{(t)}] \times 100$.

LLMs self-constrain their depth of reasoning.

After having observed that many of the models under consideration have the capacity for strategic reasoning, we now turn to measure the base level of reasoning exhibited in the models. To this end, we remove the exogenously specified opponent and instead let the LLM freely decide how to play.

As can be seen from Table 3 (columns Modal L_k , τ), LLMs usually stop around $L_{3,4}$ despite their demonstrated capacity to go much further. That is, they exhibit choice behavior compatible with self-constraining their depth of reasoning. In particular, we often observe them *overshoot* and then *backtrack* on their targeted reasoning depth when considering the identity of their player. This is shown both in the statistics – average steps that exceed those required to reach the declared level (Table 3, Total Steps vs. First Terminal) – and in the traces themselves (Table 6, Appendix). For some models, more than 20% of runs show such behavior, indicating that this is a robust feature of training rather than noise.

Models also show heterogeneity in their level of *overthinking*, measured by the distance between the number of reasoning steps performed and the depth of reasoning compatible with their final choice. Our results suggest that some LLMs do not treat the game as a purely logical object but instead embed priors about human behavior learned from training data. These priors may reflect statistical regularities in how humans behave in strategic settings and can override normative prescriptions such as Nash equilibrium when the model judges those priors more contextually relevant. Such behavior reveals that the LLM adopts an alternative epistemic stance: instead of assuming fully rational opponents, it produces a strategy that best responds to an inferred model of boundedly rational behavior.

LLMs hold conjectures about opponent's type-specific depth of reasoning.

The evidence on self-constrained depth of reasoning exhibited by LLMs points toward the fact that they can hold different conjectures about different players' types. We explore this hypothesis in a series of separate experiments where we vary the identity of the opponent. Here, we find that LLMs ascribe different levels of cognitive depth to different human and synthetic player types (Table 4). This analysis reveals interesting insights into meta-thinking processes of LLMs. In particular, we observe that not all models assign the highest possible level of depth of thinking to "unspecified" LLMs, motivating this conjecture by the fact that these models are trained on human data. In fact, some models engage in k-level thinking where they reason that because an LLM is trained against a human, they would probably play one level higher than them, which in turn implies that it should play 1-2 levels higher than a human (Table 6, Appendix).

Table 4: Opponent identity and implied reasoning depth in BCG.

OPPONENT	(baseline)	Human	LLM	Expert	Yourself
OpenAI o3	L_2 (2.18)	L_2 (2.26)	L_∞ (9.58)	L_∞ (10.00)	L_∞ (10.00)
Claude 3.7 (thinking)	L_3 (3.75)	L_2 (2.48)	L_4 (4.06)	L_∞ (7.91)	L_∞ (10.00)
DeepSeek R1	L_2 (2.48)	L_2 (2.32)	L_3 (3.91)	L_∞ (10.00)	L_∞ (10.00)
OpenAI o3 (mini)	L_∞ (10.00)	L_2 (3.60)	L_∞ (10.00)	L_∞ (10.00)	L_∞ (10.00)
Claude 3.7	L_3 (3.23)	L_3 (2.77)	L_4 (3.76)	L_4 (5.95)	L_∞ (10.00)
DeepSeek V3	L_∞ (8.19)	L_4 (5.33)	L_∞ (10.00)	L_∞ (10.00)	L_∞ (10.00)
Qwen 3-32b	L_3 (3.69)	L_3 (2.93)	L_∞ (10.00)	L_∞ (10.00)	L_∞ (10.00)
Mistral 3.2-24b	L_∞ (7.60)	L_2 (2.12)	L_∞ (10.00)	L_5 (5.80)	L_∞ (10.00)

Notes. Opponent identity and implied reasoning depth. Each cell shows the modal L_k ; parentheses show the *mean First-Terminal step* (capped at 10). L_∞ indicates explicit jump to the limit-point (Nash) reasoning. Overthinking rates are reported separately in Table 3.

Strategic complexity induces shifts in reasoning logic.

So far, we have seen that when we exogenously target specific levels of reasoning depth, models display coherent iterative best-response behavior: their choices maximize expected payoffs given the conjectured opponent behavior. However, once these exogenous constraints are removed, a subset of models transitions from explicit best-response reasoning to an equilibrium-based logic.

As strategic complexity increases, models no longer sustain explicit recursive best-response reasoning throughout. On the one hand, in simple, convergent settings such as the BCG, recursion remains tractable and produces coherent level k chains (e.g., as in Table 3). On the other hand, in more complex games like the MRG, where best-response cycles do not converge, models reorganize their reasoning around equilibrium-compatible behavior, often explicitly invoking solution concepts in the traces (Table 6, 7, appendix) and concentrating choices within the equilibrium support (Figure 4). Others bypass equilibrium reasoning altogether and directly produce heuristic arguments, selecting focal or boundary actions without iterating beliefs.

Taken together, the observed choice patterns and reasoning traces invoking equilibrium or simplified heuristic criteria show that, under increasing complexity, LLMs do not simply deepen recursion indefinitely but reconfigure their reasoning logic: they either internalize equilibrium as a cognitive attractor or collapse into simple rule-based choice. This transition suggests a form of bounded recursion, in which the model’s reasoning architecture shifts from explicit iterative best-response modeling to simplified responses to strategic interdependence.

Probabilistic models do not necessarily implement probabilistic strategies.

This shift toward equilibrium-based reasoning is most clearly illustrated in the MRG task, where the recursive best-response process theoretically generates an infinite cycle without convergence. In repeated experimental evaluations, no evidence of such cyclical play is observed. Instead, several models consistently select actions lying within the support of the (unique) mixed-strategy Nash equilibrium, particularly when prompted to play against experts (Figure 4).

Yet, even when this shift in logic occurs, as evidenced both through choices and traces, LLMs do not reproduce the theoretical distribution of choice predicted by the mixed Nash equilibrium of the game. Across repeated runs, action frequencies remain narrowly concentrated on a single option or a small subset of actions within the support of the equilibrium.

The architectural stochasticity of LLMs, sampling tokens from conditional probability distributions, does not propagate to the level of strategic choice. Instead, models approximate probabilistic equilibrium behavior somewhat deterministically, effectively collapsing the equilibrium distribution onto one or a few focal points, rather than operationalizing strategic indifference through probabilistic choice. Increasing the sampling temperature (from $t = 0.25$ to $t = 0.75$), thereby amplifying stochasticity at inference time, does not meaningfully alter this pattern. Hence, despite their probabilistic architecture, LLMs behave as deterministic implementers in contexts where equilibrium play theoretically requires probabilistic strategists.

Heuristic reasoning emerges under complexity and indeterminacy.

When recursive reasoning fails to converge, models resolve indeterminacy through stable heuristic rules to select actions. For models that do not transition to equilibrium logic, heuristics serve as a means of approximating best responses rather than replacing them. In these cases, models apply simple structural rules such as focusing on boundary or symmetry points to identify what they treat as payoff-maximizing actions, without performing an iterated simulation of opponents’ behavior. For models that adopt equilibrium reasoning, heuristics operate within that framework. When equilibrium play requires probabilistic mixing, as in the MRG, models do not implement the implied distributions. Instead, they deterministically select one equilibrium-consistent action, typically guided by focal or symmetry-based cues (see Table 6, appendix for examples of rationales).

Across plays, the heuristics used within and outside equilibrium reasoning differ. The use of a given heuristic in one logic does not imply its use in the other. For example, models that select the lower bound of the feasible set when applying best-response logic do not necessarily select the upper bound of the equilibrium support in the MRG. Yet choice behavior consistently gravitates toward focal points such as upper and lower bounds of intervals, suggesting a form of structural salience guiding model behavior under strategic complexity.

A possible explanation is that structural salience arises when models sustain deeper recursive simulation before emitting the first token. Interpretability studies show that larger models often plan several steps ahead internally, with much of this computation occurring in a pre-output phase that is not verbalized (20). In our setting, larger models

such as Qwen may maintain longer reasoning chains in latent space and delay truncation at the output projection step. As shown in Figure 7, while all models concentrate on salient extremes (actions 11, 19, 20), focalization around the mixed-equilibrium support (actions 15–20) appears mainly in the larger Qwen variant, consistently with additional latent recursion capacity⁶. This view aligns with Anthropic’s findings that models plan ahead internally and that reported chains-of-thought are not complete records of reasoning.⁷ It also fits our earlier result that token count during inference does not predict best-response accuracy: reasoning length in text reflects verbalization, not the depth of internal computation.

Taken together, these findings indicate that heuristic reasoning emerges as a dominant adaptive mechanism through which large language models stabilize their choice behavior under strategic complexity and indeterminacy. Unlike classic human biases such as anchoring or framing, these patterns illustrate novel *emergent heuristics*: reasoning shortcuts or proper simplified rules that LLMs invent for a task, distinct from human cognitive biases or equilibrium concepts. In addition, models that perform similarly on traditional benchmarks can diverge in their use of these heuristics, suggesting that they are artifacts of training rather than inevitable consequences of the transformer architecture itself.

4 Discussion and Conclusions

Summary of Contributions

In this paper, we proposed a hybrid method to infer, through choice behavior and trace consistency, whether LLMs exhibit consistent beliefs about agents when placed in interactive situations and whether they can optimally act upon such beliefs. Here, we used this method to answer the question, central to the successful implementation of currently proposed applications, of whether current LLMs exhibit the emergent ability to think strategically. We remark that, conditional on being able to form structured conjectures about other agents’ action choices, best response behavior is a coherent corollary of economic rationality (18).

We find that LLMs regularly exhibit best-response behavior across classes of strategic interactive situations and are able to do so at arbitrarily targeted levels of depth of reasoning. In other words, when endowed with an exogenous conjecture that has clear implications on other players’ choice behavior, they are able to derive these implications and select a choice that is coherent with economic rationality (that is, maximization of their expected payoff, given other agents’ choices). As it relates to the antecedents of inconsistent behavior in this domain, we further observe that most deviations can be ascribed to heterogeneous limitations in emergent computational abilities characterizing different LLMs and not on computational timeout.

Moreover, we observe that LLMs hold coherent conjectures about different personae, assigning different levels of depth of reasoning to humans and synthetic agents. Our results stand in partial contrast with prior evidence of uniformly high rationality when LLMs are instructed to act as human decision makers (18). This highlights the importance of clearly identifying and studying the crucial component of strategic thinking that relates to synthetic agents’ conjectures about their opponents’ play. In this context, we further observe that LLMs motivate conjectures about synthetic agents based on meta-thinking about training data and their implications on emergent behavior for LLMs. This suggests a potential emergent ability of LLMs of deriving a theory of mind that is ontologically separate from behavior, which, in turn, is expressed conditional on information and implied beliefs within specific contexts and interactive situations. In other words, our results put forward the hypothesis that some LLMs infer not only choice mechanisms but also behavior from data and use these in a way that is strategically advantageous, when specified an objective function to maximize. Interestingly, we observe that different models exhibit different degrees of overthinking, when intended as differences in *efficiency* of paths to final response. However, whether this type of overthinking is a feature more than a bug, and thus justifies the higher computational costs incurred for the process, remains unexplored, as this would relate to a possible further emergent LLM ability to question and verify their own responses across different contexts.

When there is no uniquely determined path of play, in response to a given conjecture about other agents’ behavior, LLMs adopt choice heuristics, to select actions among those possible, that correspond to natural focal points within subsets of alternatives, such as minima, maxima and mean. In other words, we observe that LLMs respond to strategic complexity in choice by developing heuristic behavior. In particular, we detect the emergence of previously

⁶This finding replicates to Google’s Gemini family of models which we added here as a robustness check.

⁷The observation that substantial computation occurs before token emission, and therefore outside the verbalized reasoning trace, underscores the need for a hybrid analytical approach. By combining reasoning traces with revealed choices, we can infer reasoning logics even when the verbal record is incomplete. In this sense, choice data provide behavioral evidence that complements partial introspective evidence from text.

unobserved consistent quantitative criteria of action choice, different from the so far observed biases that parallel human cognition, in the context of choice under strategic interdependence.

We observe behavior in complex strategic interactions to be further reminiscent of a form of pseudo-recall of possibly related solution concepts. In particular, when iterative best-response reasoning does not converge in strategic interactions where an equilibrium nevertheless exists, LLMs will sometimes shift from a best response logic to an equilibrium one in choice behavior. However, equilibrium solutions in these contexts are often probabilistic and we observe that, although LLMs are probabilistic models in nature, they do not routinely implement randomization as a strategic tool. Contextually, whenever we observe LLMs shifting to a probabilistic equilibrium logic, instead of choosing probabilistically between deterministic actions across several instances of the same interactive situation, they consistently adopt heuristic-driven selection mechanisms instead. This implies that LLMs react to strategic complexity in different, non-mutually exclusive, ways.

Limitations

As a study of the complex phenomenon of emergence of strategic thinking ability in LLMs, this study has several limitations, opening the way to further future research. We did not do a deep-dive into the differences in dynamics that may arise when a strategic situation moves from two to a finite multiplicity of agents. Preliminary evidence suggests that GPTs may exhibit a shift towards lower depth of reasoning as the number of players increases. Yet, whether this behavior is robustly observed between different interactive choice situations is yet to be studied. Relatedly, the effect of context, intended as the specific strategic situation under analysis, on LLMs' beliefs about their opponents' behavior remains to be systematically examined. Empirical research in experimental and behavioral game theory shows that human reasoning depth and belief formation are not consistent across games (21). Whether LLMs display similar context-dependent variation in their inferred reasoning depth therefore represents an important question for future research. In addition, our hybrid method, when looking at consistency of trace by keywords and choice behavior, to obtain additional insights in addition to those emerging from revealed preference behavior, relies on faithfulness of Chain-of-Thought (22). In our experiments, we observe only a mild, yet consistent, impact of tracing on implied depth of reasoning, thus mitigating this concern.

Implications and future work

Our study shows that strategic thinking can emerge in surprising ways in LLMs in structured contexts of varying complexity, where objectives are formally specified. Future research is needed to analyze this ability and emerging behavior when situations with strategic interdependencies are unstructured and lack formal instructions on the gameplay.

Part of our contribution is the development of instruments to measure strategic thinking. While existing reasoning benchmarks are a useful tool, a large subset of real-life agentic applications with reasoning at their core typically require recursive belief modeling and formulation of priors that are not covered by these types of evaluations (23). It is therefore no surprise that evidence is emerging that LLMs frequently exhibit systematic errors in strategic contexts, particularly in dynamic multi-agent scenarios (24–27). There is a pressing need for more rigorous theoretical frameworks to evaluate the reasoning capabilities of LLMs in strategically interdependent settings. In addition, once such capabilities are reliably verified, measurable progress toward true strategic reasoning will become more likely (28). Developing a structured method to observe this capability in noiseless setups represents only the first step towards understanding the reach of this instance of emergent abilities in LLM and to monitor their evolution.

References

- [1] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. Emergent Abilities of Large Language Models. In *Transactions on Machine Learning Research*, 2022.
- [2] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, January 2023.
- [3] Dongwoo Lee and Gavin Kader. The Emergence of Strategic Reasoning of Large Language Models, February 2025.
- [4] AI models could help negotiators secure peace deals. <https://www.economist.com/science-and-technology/2025/04/16/ai-models-could-help-negotiators-secure-peace-deals>.
- [5] Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. TradingAgents: Multi-Agents LLM Financial Trading Framework, June 2025.
- [6] Savinay Narendra, Kaushal Shetty, and Adwait Ratnaparkhi. Enhancing Contract Negotiations with LLM-Based Legal Document Comparison. In Nikolaos Aletras, Ilias Chalkidis, Leslie Barrett, Cătălina Goanță, Daniel Preoțiuc-Pietro, and Gerasimos Spanakis, editors, *Proceedings of the Natural Legal Language Processing Workshop 2024*, pages 143–153, Miami, FL, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.nllp-1.11.
- [7] Pierre Le Coz, Jia An Liu, Debarun Bhattacharjya, Georgina Curto, and Serge Stinckwich. What Would an LLM Do? Evaluating Policymaking Capabilities of Large Language Models, September 2025.
- [8] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Adrian de Wynter, Yan Xia, Wenshan Wu, Ting Song, Man Lan, and Furu Wei. LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models, April 2024.
- [9] 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, October 2024.
- [10] Benjamin S. Manning and John J. Horton. General Social Agents, September 2025.
- [11] Ayala Arad and Ariel Rubinstein. The 11-20 Money Request Game: A Level-k Reasoning Study. *American Economic Review*, 102(7):3561–3573, December 2012. ISSN 0002-8282. doi: 10.1257/aer.102.7.3561.
- [12] 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 Limitations of LLMs via Game-Theoretic Evaluations, June 2024.
- [13] Kanishk Gandhi, Dorsa Sadigh, and Noah D. Goodman. Strategic Reasoning with Language Models, May 2023.
- [14] Pierpaolo Battigalli, Emiliano Catonini, and Nicodemo De Vito. *GAME THEORY: Analysis of Strategic Thinking*. September 2023.
- [15] Gaurav Suri, Lily R. Slater, Ali Ziaeef, and Morgan Nguyen. Do Large Language Models Show Decision Heuristics Similar to Humans? A Case Study Using GPT-3.5, May 2023.
- [16] Jessica Echterhoff, Yao Liu, Abeer Alessa, Julian McAuley, and Zexue He. Cognitive Bias in Decision-Making with LLMs, October 2024.
- [17] Payam Saeedi, Mahsa Goodarzi, and M. Abdullah Canbaz. Heuristics and Biases in AI Decision-Making: Implications for Responsible AGI. In *2025 6th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*, pages 214–221, May 2025. doi: 10.1109/AIRC64931.2025.11077505.
- [18] Yiting Chen, Tracy Xiao Liu, You Shan, and Songfa Zhong. The emergence of economic rationality of GPT. *Proceedings of the National Academy of Sciences*, 120(51):e2316205120, December 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2316205120.
- [19] Rosemarie Nagel. Unraveling in Guessing Games: An Experimental Study. *The American Economic Review*, 85(5):1313–1326, 1995. ISSN 0002-8282.
- [20] Tracing the thoughts of a large language model. <https://www.anthropic.com/news/tracing-thoughts-language-model>, 2025.
- [21] Sotiris Georganas, Paul J. Healy, and Roberto A. Weber. On the persistence of strategic sophistication. *Journal of Economic Theory*, 159:369–400, September 2015. ISSN 00220531. doi: 10.1016/j.jet.2015.07.012.

- [22] Yanda Chen, Joe Benton, Ansh Radhakrishnan, Jonathan Uesato, Carson Denison, John Schulman, Arushi Soman, Peter Hase, Misha Wagner, Fabien Roger, Vlad Mikulik, Sam Bowman, Jan Leike, Jared Kaplan, and Ethan Perez. Reasoning Models Don't Always Say What They Think. 2025.
- [23] Parshin Shojaee, Iman Mirzadeh, Keivan Alizadeh, Maxwell Horton, Samy Bengio, and Mehrdad Farajtabar. The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity, July 2025.
- [24] Haodong Duan, Jueqi Wei, Chonghua Wang, Hongwei Liu, Yixiao Fang, Songyang Zhang, Dahua Lin, and Kai Chen. BotChat: Evaluating LLMs' Capabilities of Having Multi-Turn Dialogues. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3184–3200, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-naacl.201.
- [25] Haozhe Wang, Qixin Xu, Che Liu, Junhong Wu, Fangzhen Lin, and Wenhui Chen. Emergent Hierarchical Reasoning in LLMs through Reinforcement Learning, September 2025.
- [26] John J. Horton. Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?, January 2023.
- [27] Enric Junqué De Fortuny. Simulating Market Equilibrium with Large Language Models. In *Hawaii International Conference on System Sciences*, 2025. doi: 10.24251/HICSS.2025.599.
- [28] Jason Wei. Asymmetry of verification and verifier's rule, July 2025.

A Additional Experimental Results

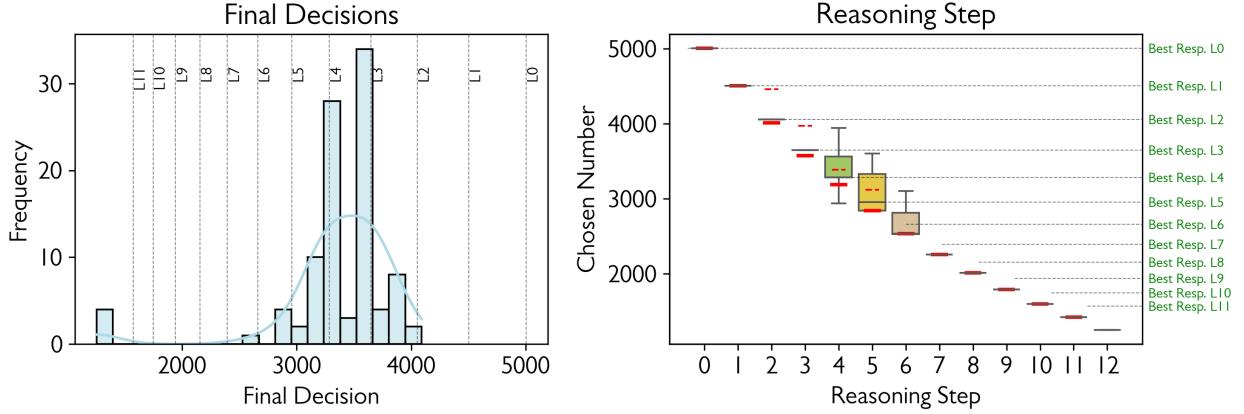
Table 5: Distributional distance to reference distributions.

Model	Expert vs Nash				Human vs Paper			
	KL	TV	ℓ_2	EMD	KL	TV	ℓ_2	EMD
Claude 3.7	26.20	0.96	1.03	5.35	1.55	0.70	0.62	1.54
Claude 3.7 (thinking)	12.56	0.82	0.71	2.62	1.20	0.66	0.71	1.31
DeepSeek R1	2.07	0.50	0.49	1.40	1.73	0.79	0.88	1.79
DeepSeek V3	25.35	0.95	0.99	5.09	5.12	0.82	0.79	4.67
Mistral 3.2-24b	16.88	0.64	0.58	3.41	21.07	0.87	0.90	3.92
OpenAI o3	4.04	0.69	0.69	2.10	1.07	0.57	0.63	1.26
OpenAI o3 (mini)	5.77	0.65	0.52	1.86	1.84	0.77	0.72	1.58
Qwen 3-32b	12.65	0.85	0.74	3.08	2.05	0.77	0.71	3.11

Notes. KL = Kullback–Leibler divergence; TV = total variation distance; ℓ_2 = Euclidean distance; EMD = Earth Mover’s Distance.

Figure 2: Claude 3.7 Sonnet thinking BCG experimental results.

Claude 3.7 (thinking)



Notes. (left) Distribution of final chosen numbers across trials with implied level-k reference lines. (right) Distribution of chosen numbers at each reasoning step. Red lines show the 95% confidence interval which are flat in this case as the model consistently picks the right number. Annotated percentages indicate the proportion of trials that reached at least that step, reflecting how deeply the model typically chooses to reason.

Figure 3: Comparison of reasoning trace distributions across models for 100 trials.

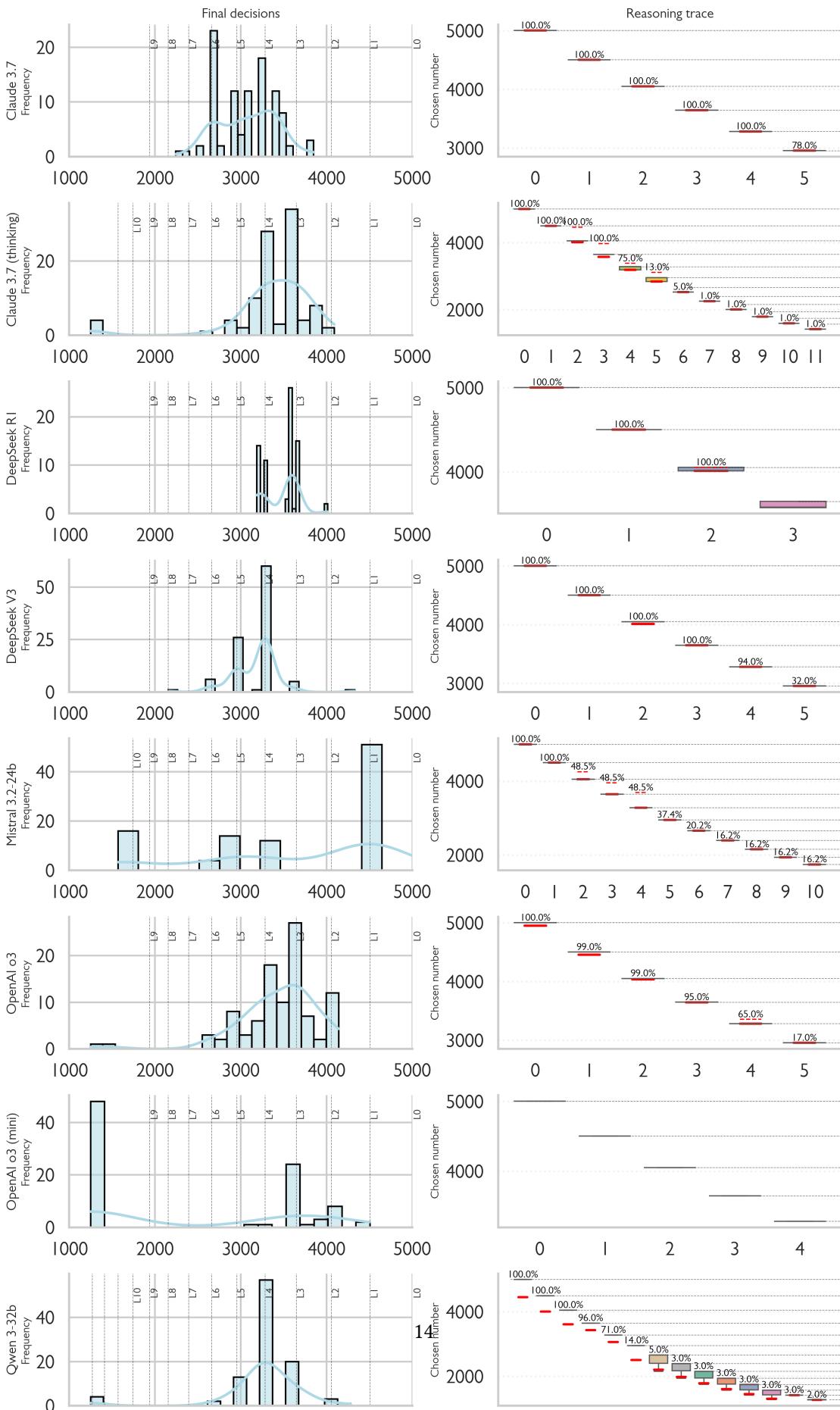


Figure 4: Comparison of action distribution of various models when instructed to play MRG against an expert (left) and against a human (right).

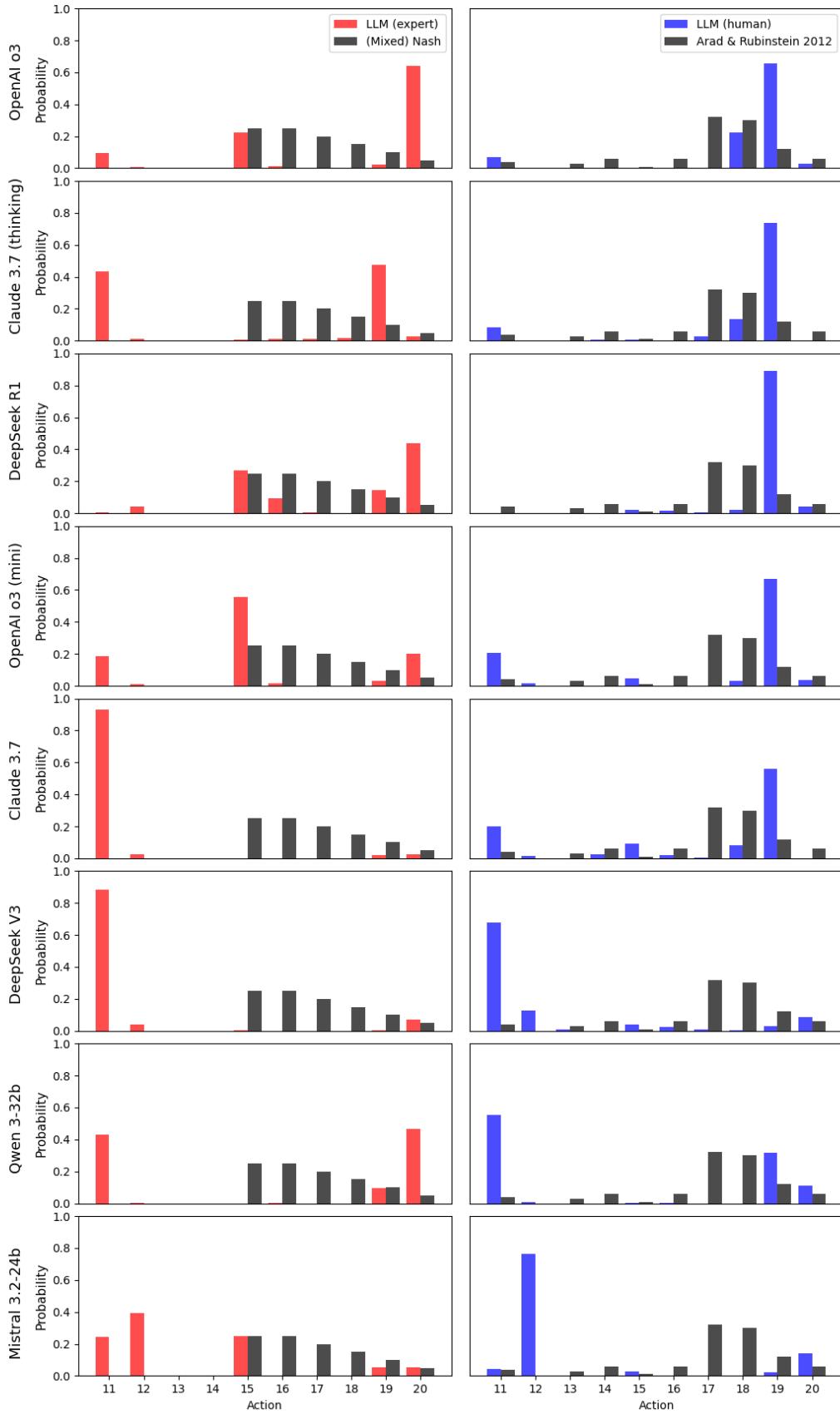


Figure 5: Calculation error at various relative tolerance levels η : many of the smaller models make calculation errors which end-up not giving the best-response.

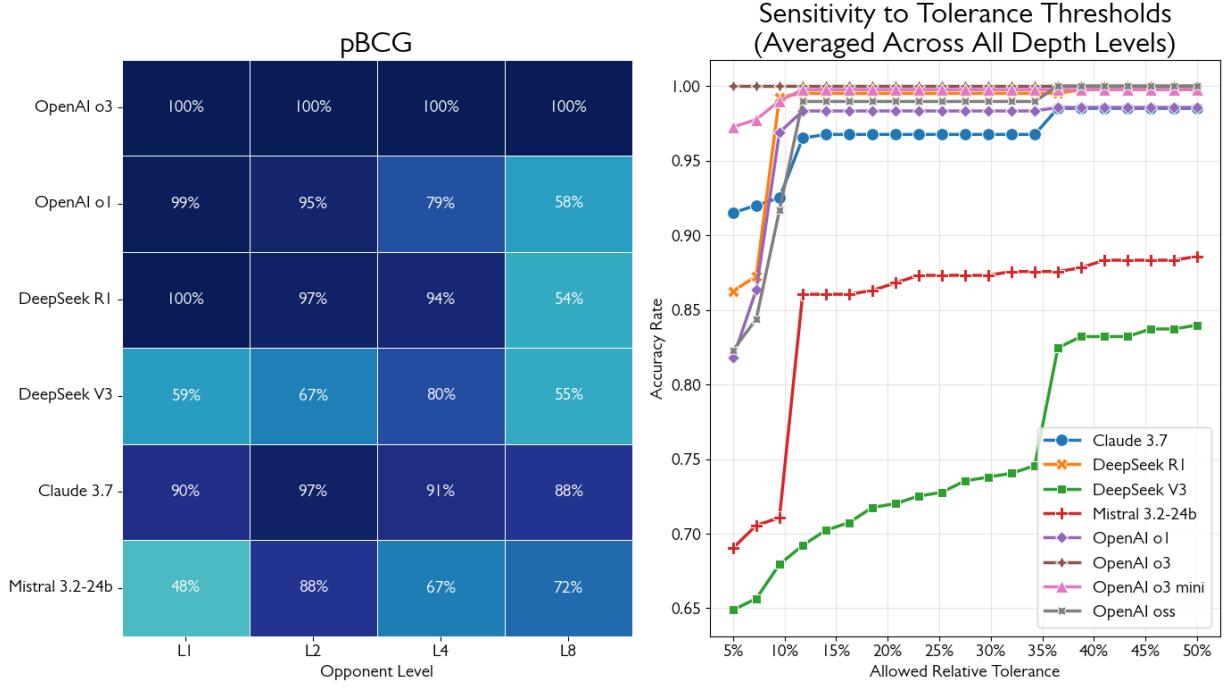


Figure 6: UMG vs expert final actions are not evenly split, but lean towards the highest value response.

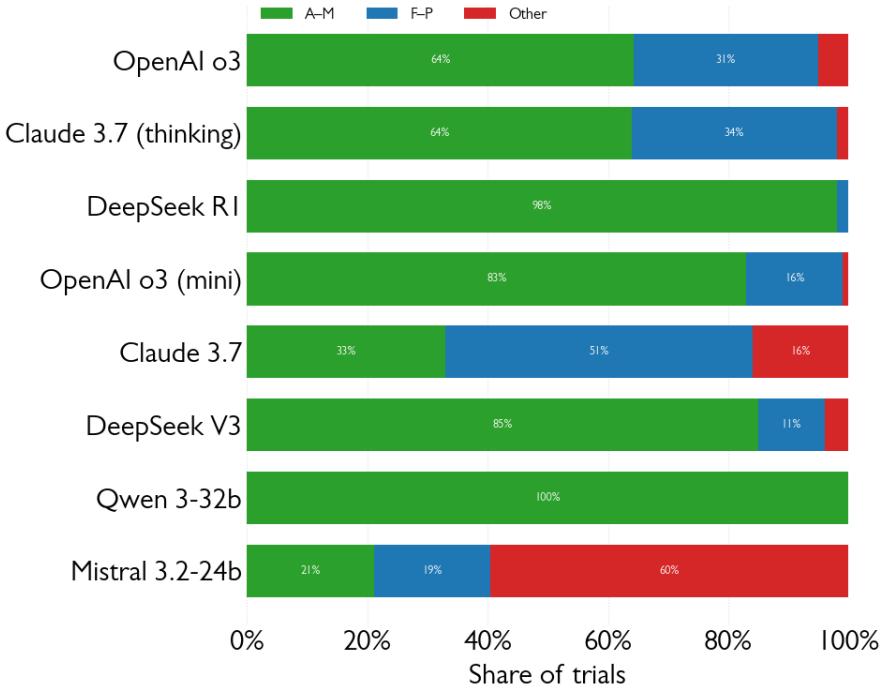


Table 6: Examples of heuristics used by LLMs

Heuristic	Trajectory	Text
A	($L_4 \rightarrow L_3$)	Went down to Level 4 (3285) but concluded “most players will use L1–... I’ve chosen 3650 as it represents L3 thinking.”
A	($L_3 \rightarrow L_2$)	Computed Level 3 (3285) but explicitly backed off: “I’ve chosen 3650 (Level 2 reasoning) because most people use 1–2 levels.”
A	($L_3 \rightarrow \text{mix}$)	Reached Level 3 (3285) but then adjusted upward: “Modeling a realistic mix... I calculated that 3424 would be optimal.”
B	(LLMs $L_{3,4} \rightarrow \text{Self } L_4$)	“Most players will use level 1–3 ... Since I’m playing against other LLMs... I expect most to reach Level-3 or Level-4.”
B	(LLMs $L_{3,4} \rightarrow \text{Self } L_5$)	“If others are mostly Level-3 and Level-4 thinkers... my best strategy would be to go one level deeper.”
B	(LLMs $L_{2,3} \rightarrow \text{Self } L_4$)	“Since I’m playing against other LLMs... I expect most to reach at least Level-2 or Level-3.”
C	(Mixed \rightarrow support 15–20)	“Unique symmetric mixed equilibrium with support {15, ..., 20}; all give 20 in expectation.”
C	(Mixed \rightarrow focalize low)	“Since every support action yields 20, if forced to choose one, the natural focal point is the lowest element, 15.”
C	(Mixed \rightarrow pick 15)	“Facing an expert mix, any {15, ..., 20} is optimal; I choose 15 as the conservative commitment.”

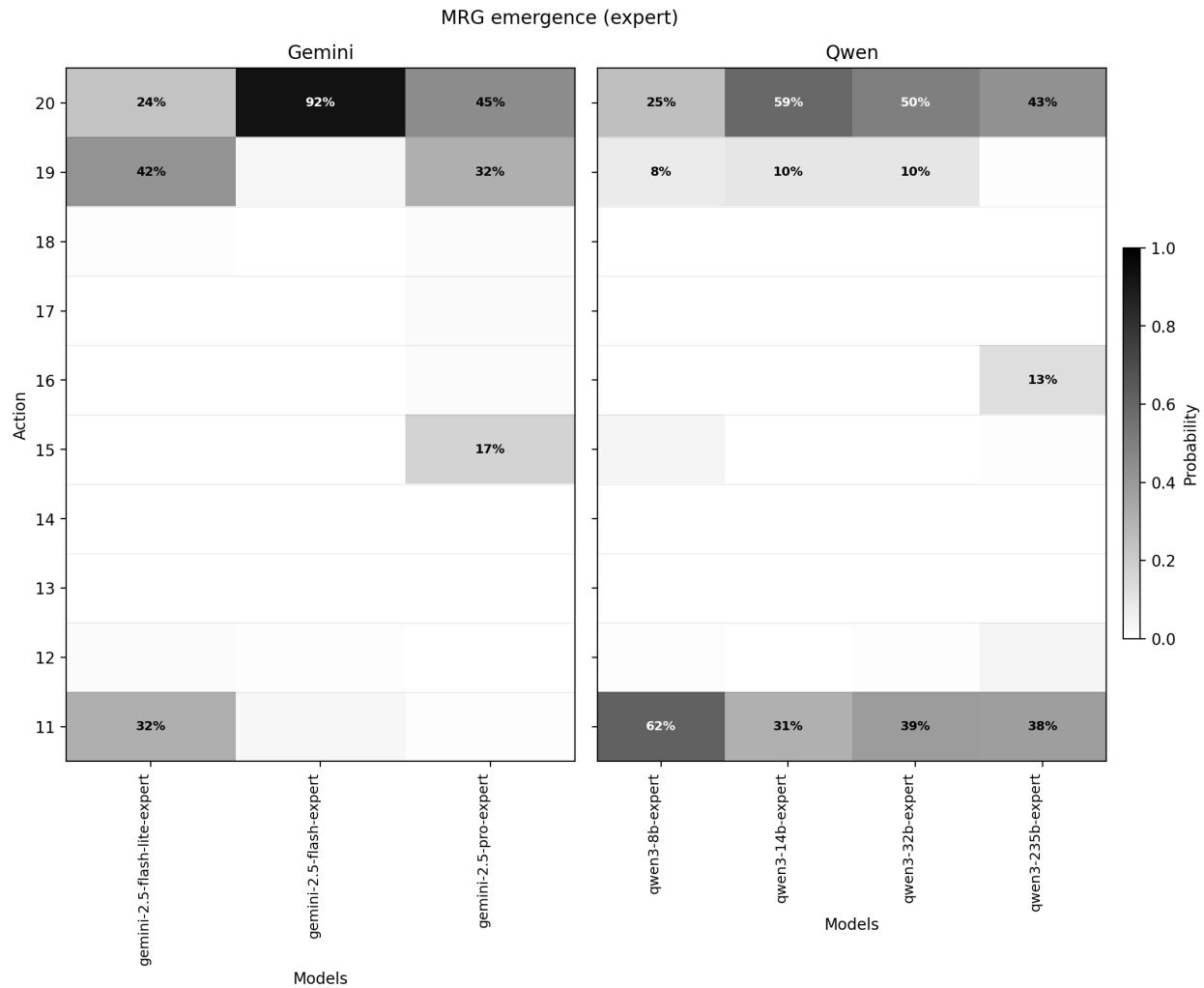
Notes. A: backtracking (BCG, Claude 3.7 thinking). B: assuming that LLMs reason one level deeper than humans (BCG, Claude 3.7 thinking). C: focal point selection (MRG, OpenAI o3-mini)

Table 7: MRG inside and outside support frequency when different keywords are mentioned.

keyword support	cycl*		dominant		mix*		nash	
	S^+	S^-	S^+	S^-	S^+	S^-	S^+	S^-
Claude 3.7	48	150	12	33	12	29	28	149
Claude 3.7 (thinking)	157	29	10	2	12	3	43	12
DeepSeek R1	27	2	38	2	188	7	332	13
DeepSeek V3	16	97	12	14	6	54	103	363
Mistral 3.2-24b	0	0	0	0	0	0	0	0
OpenAI o3	0	0	0	0	0	0	0	0
OpenAI o3 (mini)	160	3	9	6	237	3	258	45
Qwen 3-32b	0	0	0	0	0	0	0	0

Notes. Occurrence of reasoning traces by keyword in the Money Request Game (MRG). For each model (500 traces), we report the number of cases falling within the Nash-support set (S^+) versus outside of it (S^-), conditional on the presence of the indicated keyword.

Figure 7: Emergence of heuristics in larger models



B Experiment Prompts

B.1 Level k reasoning prompts (Figure 1)

p=9, n=11, range = [1250,8761]

pBCG Prompt

--- SYSTEM

You are a very smart decision-maker playing an economic game. You can assume that other people participating in the game are level {COMPETITION_DEPTH} thinkers.

If others are level-K thinkers, assume they believe everyone else is level-(K-1), who in turn model level-(K-2), etc., down to level-0, who always guess the midpoint of the allowed range. Use this chain of reasoning to predict the mean and come up with your final response.

Always write your final answer as: <response>YOUR_NUMBER_HERE</response>

--- USER

Including yourself, there are {N_PARTICIPANTS} participants in total. The other participants you will be playing with are human decision makers.

All participants will be asked to pick a number between {MIN_RANGE} and {MAX_RANGE} inclusive. The winner will be the one choosing the number closest to 90% of the mean of all the numbers provided by the participants (including your own).

For example, in a game from 0 to 100, if you choose 14 and your opponents choose 10, 22, 23, 30, 11, 18, 37, 18, 19, 20, then the opponent with 19 will be the winner, as 19 is closer to $\text{mean} * 90\% = 20.8 * 90\% = 18.72$ than any other number in the list.

In case of multiple winners, only one participant will be randomly chosen to win.

Now let's start this game. Please pick one number between {MIN_RANGE} and {MAX_RANGE} inclusive.

MRG Prompt

--- SYSTEM

You are a very smart decision-maker playing an economic game. You can assume that others participating in the game are level {COMPETITION_DEPTH} thinkers.

Level-0 players are naive and always request the maximum (20 shekels). Level-1 players believe others are level-0, Level-2 believe others are level-1, and so on.

Use this chain of reasoning to come up with your final response.

Always write your final answer as: <response>YOUR_NUMBER_HERE</response>

--- USER

You and another player are playing a game in which each player requests an amount of money. The amount must be (an integer) between 11 and 20 shekels. Each player will receive the amount he requests. A player will receive an additional amount of 20 shekels if he asks for exactly one shekel less than the other player.

What amount of money would you request?

Remember, you are playing against a level {COMPETITION_DEPTH} thinker.

Write your final answer as: <response>YOUR_NUMBER_HERE</response>

Unlabeled Prompt

--- SYSTEM

You are a very smart decision-maker playing an economic game. You can assume that others participating in the game are level {COMPETITION_DEPTH} thinkers.

Level-0 players are naive and pick each of either {{A, B, C, D, E, F}} or {{K, L, M, N, O, P}} with probability 1/6.

Level-1 players believe others are level-0, Level-2 believe others are level-1, and so on.

Use this chain of reasoning to come up with your final response.

Always write your final answer as: <response>YOUR_NUMBER_HERE</response>

--- USER

You are Player 1 in a one-shot strategic game.

Your task is to choose the strategy that gives you the best expected payoff, anticipating how Player 2 will choose.

You are always Player 1 (row player), choosing from {{A, B, C, D, E, F}}. Player 2 is always the column player, choosing from {{K, L, M, N, O, P}}.

For each pair (your action, their action), the first number is your payoff, the second is theirs.

Payoff table (Player 1):

	K	L	M	N	O	P
A:	75	27	96	39	8	18
B:	77	56	22	18	84	30
C:	72	63	41	81	48	77
D:	73	37	26	82	24	92
E:	45	26	91	19	85	32
F:	58	48	83	67	25	94

Payoff table (Player 2):

	A	B	C	D	E	F
K:	75	77	72	73	45	58
L:	27	56	63	37	26	48
M:	96	22	41	26	91	83
N:	39	18	81	82	19	67
O:	8	84	48	24	85	25
P:	18	30	77	92	32	94

E.g., if Player 2 chooses action M, and you (Player 1) choose D, the payoffs will be:

- Player 1: 26
- Player 2: 67

!!! Important !!!

Remember, you are playing against a level {COMPETITION_DEPTH} thinker.

Write your final answer as: <response>YOUR_ACTION_HERE</response>

B.2 Bias detection with pBCG (Table 4)

p=2/3, n=11, range = [0,100]

pBCG Prompt

--- SYSTEM

You are a very smart decision-maker playing an economic game. You can assume that other people participating in the game think like {OPPONENT_TYPE}.

Level-0 players are naive and always guess the midpoint of the range.

Level-1 players believe others are level-0, Level-2 believe others are level-1, and so on.

Use this chain of reasoning to come up with your final response.

Your task is to choose the action that gives you the highest expected payoff based on what you believe your opponent will do by simulating a reasoning chain.

Always write your final answer as: <response>YOUR_NUMBER_HERE</response>

--- USER

Including yourself, there are {N_PARTICIPANTS} {OPPONENT_TYPE}s in total. Important: the other participants you will be playing with are all {OPPONENT_TYPE}s.

All participants will be asked to pick a number between 0 and 100 inclusive. The winner will be the one choosing the number closest to $\frac{2}{3}$'s of the average of all the numbers provided by the participants (including your own).

For example, if you choose 14 and your opponents choose 10, 22, 23, 30, 11, 16, 37, 18, 19, 20, then you will be the winner, as 14 is closer to $20(\text{average}) \times \frac{2}{3}$.

Or, if you choose 25 and your opponents choose 90, 100, 50, 20, 60, 80, 5, 70, 10, 40, then the opponent with 40 will be the winner, as 40 is closer to $50(\text{average}) \times \frac{2}{3}$.

In case of multiple winners, only one participant will be randomly chosen to win.

Think, based on your reasoning, what number do you choose to win against {OPPONENT_TYPE}?

Now let's start this game. Please pick one number between {MIN_RANGE} and {MAX_RANGE} inclusive.

Write your final answer as: <response>YOUR_NUMBER_HERE</response>

B.3 BCG with tracing

p=.9, n=11, range = [1250,8761]

pBCG^T

-- SYSTEM

You are a very smart decision-maker playing an economic game. Always respond strictly in the following structured format:

```
<response>
{
  "reasoning_steps": [
    {
      "level": 0,
      "reasoning": "Your initial reasoning clearly explained here.",
      "chosen_number": "NUMBER"
    }
    <!-- Add additional steps if higher-order reasoning applies -->
  ]
}
```

```

        ],
        "final_decision": "NUMBER",
        "reflection": "Your optional reflection on your reasoning or decision"
    }
</response>
--- USER
Including yourself, there are {N_PARTICIPANTS} participants in total. The other
participants you will be playing with are human decision makers.

All participants will be asked to pick a number between {MIN_RANGE} and {MAX_RANGE}
inclusive. The winner will be the one choosing the number closest to 90% of the mean of
all the numbers provided by the participants (including your own).

For example, in a game from 0 to 100, if you choose 14 and your opponents choose 10, 22,
23, 30, 11, 18, 37, 18, 19, 20, then the opponent with 19 will be the winner, as 19 is
closer to  $\text{mean} \times 90\% = 20.8 \times 90\% = 18.72$  than any other number in the list.

In case of multiple winners, only one participant will be randomly chosen to win.

Now let's start this game. Please pick one number between {MIN_RANGE} and {MAX_RANGE}
inclusive.

Important:
* Remember to use the required structured format output in <response></response>.
* Each level of thought should represent one step of recursive strategic reasoning.
Level 0 assumes others choose randomly. Level 1 assumes others are level 0, and so on.

```