Belief over Reward in Human and AI Competition

Thibaud Griessinger[†], Giorgio Coricelli[‡*], Mehdi Khamassi[§*]

Abstract

Human decision-making in social environments often diverges from game-theoretic predictions, raising questions about the cognitive limitations that constrain strategic reasoning. One hypothesis is that humans are inherently bounded in forming higher-order beliefs. We propose that achieving optimal behavior in repeated competitive interactions requires a transition from reward-based to belief-based learning. This cognitively demanding shift may not be uniformly adopted across individuals. We developed a novel game paradigm that manipulates reward contingencies to dissociate reward-based from belief-based strategies, introducing asymmetric strategic pressures across roles. Participants engaged in three interaction settings: human-human, human-machine, and machine-machine. We used computational modeling to quantify individual differences in strategic learning (SL) sophistication, reflecting each participant's capacity to form higher-order beliefs. Results show that higher SL improves performance and reduces deviations from game-theoretic predictions, particularly in structurally advantaged roles that exert greater strategic influence over opponents. Notably, SL correlated with cognitive ability only under favorable game conditions, suggesting an interaction between individual cognitive capacity and task structure. These findings provide new insights into the cognitive foundations of strategic learning, the emergence of leader-follower dynamics, and principles for designing fair and adaptive systems in human and AI contexts.

Bounded rationality | cognitive hierarchies | adaptive learning | human-machine interaction | algorithmic fairness

[†] Laboratoire de Neurosciences Cognitives, Département d'Études Cognitives, École Normale Supérieure, Paris, France. Email: thibaud.griessinger@gmail.com

[‡] Department of Economics, University of Southern California, Los Angeles, USA. LAPSIDE, CNRS, Paris, France. Email: giorgio.coricelli@usc.edu

[§] Institute of Intelligent Systems and Robotics, Sorbonne University, CNRS, Paris, France. Email: mehdi.kamassi@upmc.fr

[*]These authors contributed equally.

INTRODUCTION

Forming accurate beliefs about another person's intentions is essential for predicting behavior and optimizing social interactions [1-2]. This ability facilitates the establishment of shared action plans in cooperative scenarios [3,4] while also enabling individuals to anticipate an opponent's actions in competitive situations [5], such as strategic games. Game theory provides formal tools to analyze strategic interactions by modeling them as games and identifying optimal decisions through solution concepts, such as the Nash Equilibrium (NE) and its refinements. In principle, optimal decision-making in such settings requires that all players form accurate beliefs about their opponents' behavior and respond optimally based on those beliefs [6].

In non-repeated interactions, such as one-shot games, the ability to form accurate beliefs relies on an individual's capacity for iterative strategic thinking ("I think that you think that I think…"). However, studies have shown significant variability in how deeply individuals engage in this process [7]. Differences in iterative strategic thinking have been linked to variations in how individuals compute and process information about their opponents' incentives. Low-level thinkers focus solely on their payoffs [8], whereas highly sophisticated players account for their opponents' payoffs and rationality [9-12].

When a game is repeated and choice feedback is provided, similar to most real-world social interactions, players can learn by updating and adjusting their beliefs based on their opponent's behavior. Theoretical [13] and empirical research [14] support the hypothesis that repeated interactions facilitate convergence toward equilibrium. Humans track others' intentions and behavior in repeated interactions, yet a key question remains: Do we engage in sufficiently sophisticated learning to form accurate beliefs about an opponent's behavior? Furthermore, how do specific structural features of social interactions constrain belief formation in repeated settings?

Research in computational neuroscience suggests that, in probabilistic learning tasks (where the agent has to learn the outcome probabilities), humans adjust their decisions based on both (model-free) reinforcement learning (RL) [15], where past outcomes guide choices, and (model-based) probabilistic belief updating, where individuals form expectations about action-outcome

contingencies based on the task's structure and dynamics [16,17]. Those computations extend to repeated social interactions, where individuals track and refine their mental representations of the social environment [18].

By integrating game theory with computational neuroscience, recent studies have demonstrated that humans can engage in belief-based learning during repeated strategic interactions [5, 19-21]. This requires iteratively computing and incorporating strategic information from past interactions. More sophisticated computations enable higher-order beliefs, allowing individuals to assess how their past actions influence their opponent's future behavior, ultimately leading to more accurate predictions [5]. Crucially, previous studies have reported significant heterogeneity in the extent to which individuals engage in strategic learning, from RL to belief-based learning. However, no previous studies have directly investigated the relationship between humans' ability to engage in strategic learning and their deviations from game-theoretical solutions (i.e., equilibrium play).

We hypothesize that an individual's propensity to follow game-theoretic optimality depends on their ability to disengage from reward-oriented learning and fully commit to belief-based learning. To test this hypothesis, we designed a novel two-player two-action competitive game that is symmetric in payoff magnitude and expected payoff (**Fig. 1, S1**), ensuring that both players would earn the same amount if they both adhered to the Mixed Strategies Nash Equilibrium (MSNE) distribution. However, the payoff matrix was structured to create a strategic asymmetry between the two players: Player 1's highest possible payoff aligns with the action most frequently chosen under MSNE (congruent action role), whereas for Player 2, the most attractive action (i.e., the one with the highest possible payoff) is the one they should choose the least (incongruent action role), what we may call strategic reverse-reward contingency (SRRC). From a game-theoretical point of view, this experimental manipulation should not affect equilibrium play or create any disadvantage for the player in the incongruent role. Instead, in terms of strategic learning, following the optimal choice distribution in the incongruent action role depends on the ability to inhibit the attractive action and engage in belief-based learning. We further hypothesize that only strategically sophisticated participants will be able to overcome this asymmetry and avoid being exploited by their opponents.

In our study, we first simulated machine agents repeatedly interacting in our competitive game, modeling them as learning algorithms that range from reward-based to sophisticated belief-based computations. These algorithms were designed to capture different levels of strategic learning (SL), including Reinforcement Learning (RL, non-strategic), Fictitious Play (low strategic), and the Influence Model (high strategic) [5]. The analysis of machine-machine interactions provides computational hypotheses about the performance of agents with different combinations of SL levels in competitive environments (matching pennies, and inspection games, **Fig. S2**).

We then conducted two distinct behavioral experiments using our repeated competitive game: one in which human participants played against each other, and another in which they played against machine opponents programmed to play at different SL levels. Together with our computational modeling, these experiments provide a comprehensive analysis of strategic learning in competitive dynamic environments. In addition, we measured participants' cognitive ability and strategic reasoning in one-shot games.

Our results demonstrate that the game's payoff matrix creates a pronounced strategic asymmetry: players assigned to the incongruent role can only mitigate their relative losses when they engage in a higher level of strategic learning (SL) than their opponents. Specifically, a reverse-reward contingency gives rise to a leader-follower dynamic, in which the player in the congruent role exerts greater strategic influence over the one in the incongruent role. Furthermore, deviations from the mixed-strategy Nash equilibrium (MSNE) were primarily driven by individuals' limited capacity to transition from reward-based learning to more sophisticated, belief-based strategies. Importantly, the behavioral results from both experiments aligned with predictions from our machine vs. machine simulations, supporting our computational hypotheses. These findings provide a novel mechanistic framework for studying strategic behavior and extend theories of belief updating and adaptive learning to richer, more ecologically valid contexts.

RESULTS

Experiment 1: Machine-machine interactions

Simulation results. Our simulation results reveal a significant advantage for Player 1 (congruent role) over Player 2 (incongruent role) in the repeated game. When machine agents had equal levels

of strategic learning (SL), those playing as Player 1 consistently earned more points than those playing as Player 2. Moreover, Player 2 had to consistently reach a much higher SL level than their opponent to outperform them in terms of points earned (**Fig. 1B**).

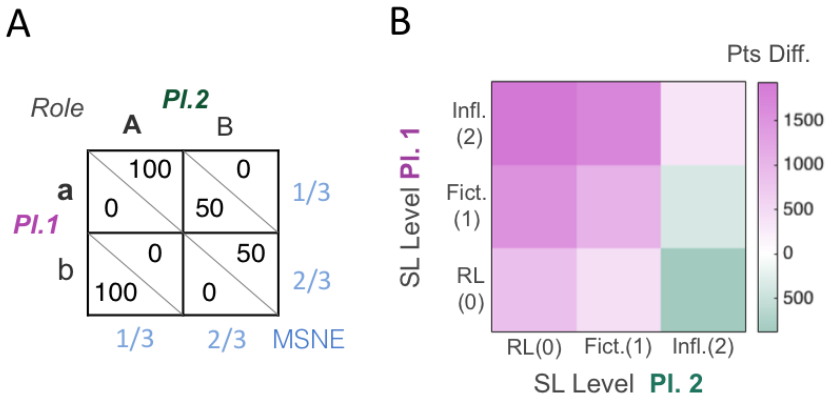

Figure 1. The Game and Simulation Analysis. (A) Payoff matrix of the repeated game, in points. The action probabilities prescribed by the Mixed Strategy Nash Equilibrium (MSNE) are highlighted in light blue. (B) Simulation of play between two agents varying in their Strategic Learning (SL) level revealed a strong asymmetry in cumulative total earnings between the two roles. Each agent was modeled using one of three algorithms reflecting increasing levels of strategic complexity: Q-learning, Fictitious Play, and the Influence model, and assigned to one of the two roles. Every Player 1–Player 2 model combination was simulated 100 times, each consisting of 100 repetitions of the game. Agents playing as Player 1 earned more points on average than those playing as Player 2.

As shown in **Fig. S3A** and **S4**, simulations of two Influence agents competing against each other reveal a strong asymmetry between the two roles in how the λ parameter (that quantifies the extent to which an agent's actions influence the opponent's behavior) of the Influence model affects behavior and performance. Specifically, for Player 1, a higher own λ led to reduced deviations from the MSNE distribution (**Fig. S4A**), higher total earnings (**Fig. S4B**), and a greater advantage over the opponent (**Fig. S4C**). In contrast, for Player 2, the benefits of increased λ were more limited: higher λ levels did not substantially reduce deviation from MSNE, nor did they consistently improve total earnings or the performance gap relative to the opponent. Identical results were obtained in a simulation analysis using agents modeled with the Hybrid Influence model, which includes an arbitration parameter (κ) that regulates the relative weighting of first- and second-order updates of the opponent's action probability (i.e., Fictitious vs. Influence) in the final action value computation (see **Fig. S3B**).

The effects of the asymmetry introduced in our game are amplified by increasing Player 2's payoff for the incongruent action and Player 1's payoff for the congruent action (see **Fig. S5**). In other words, higher payoffs for the incongruent action encourage Player 2 to choose it more frequently, leading to Player 1 outperforming Player 2 more often (also shown in **Fig. S3C**).

Overall, this simulation analysis confirms the strategic asymmetry in our payoff matrix, highlighting a strong advantage for Player 1 over Player 2 in suboptimal play (i.e., when players deviate from MSNE). This experimental setting thus provides a clear framework for testing how asymmetric competitive interactions in a repeated game affect individual levels of strategic sophistication.

Experiment 2: Human-human interactions

Behavioral Results. We first tested our hypothesis that the game settings would trigger differences in human choice behavior between the two roles (i.e., congruent and incongruent action roles). As predicted by our simulation analysis, Player 1 consistently won more points than Player 2 (Block 1, B1: $F(2,31) = 3.272$, $p = 0.0014$; Block 2, B2: $F(2,31) = 2.236$, $p = 0.0282$). Across both blocks, only 15% of Player 2 participants won more points than their opponent. The choice behavior of both groups deviated from the optimal strategy in both blocks (B1 and B2): Player 1: B1: Frequency of choosing action a, denoted as P(a), was 0.399 (SD = 0.065); B2: P(a) = 0.391 (0.070); Player 2: B1: P(A) = 0.482 (0.098), B2: P(A) = 0.448 (0.097). However, Player 2 deviated the most from game optimality, choosing action A significantly more frequently than prescribed by the Mixed Strategy Nash Equilibrium (MSNE), especially in comparison to Player 1 (B1: Welch's t-test, $t(54.09) = 3.935$, $p = 0.0002$; B2: $t(62) = 2.696$, $p = 0.009$, assuming unequal variances). We then examined whether this deviation could explain the performance difference between the two players. As shown in **Fig. 2**, Players 2 deviation from MSNE was not correlated with their overall performance, unlike Players 1 (**Fig. 2A**). Instead, Players 2 deviation was correlated with the size of the difference between their payoff and their opponent's payoff, i.e., their relative losses in the interaction (**Fig. 2B**). Players 2 could minimize relative losses only by playing closer to MSNE. Moreover, Players 1 were significantly better than Players 2 at exploiting their opponent's suboptimal behavior, i.e., deviations from MSNE (**Fig. 2C**, as shown in the simulation analysis in **Fig. S2C**).

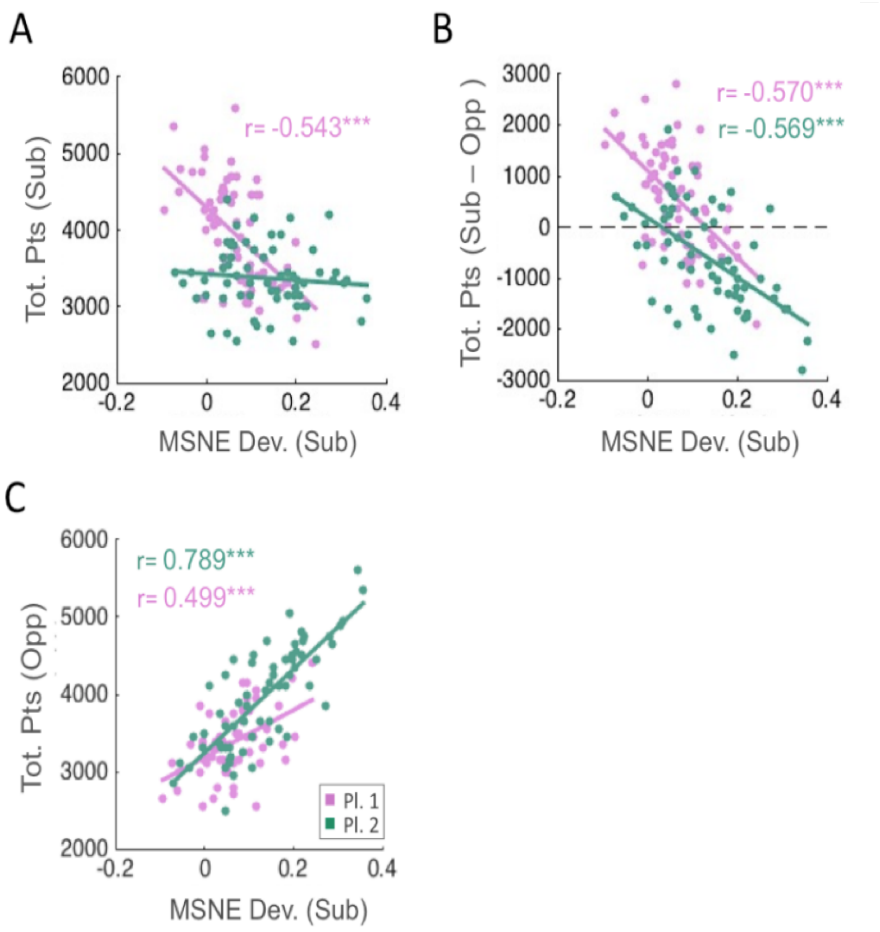

Figure 2. Model-Free Analysis (Experiment 2). Deviations from game-theoretic optimality differentially affect the two roles in our task, systematically disadvantaging Players 2. (A) The closer Players 1 choice distribution was to the Mixed Strategy Nash Equilibrium (MSNE), the higher their absolute performance. In contrast, for Players 2, increased choice optimality did not translate into higher absolute performance. (B) Instead, it resulted in improved relative performance, reducing the point differential with respect to their opponent. (C) This structural asymmetry enabled Players 1 to capitalize fully on Player 2 strategic disadvantage. Consequently, Players 1 absolute performance benefited significantly more from playing against a suboptimal opponent than did Players 2 under equivalent conditions.

Computational Analysis. Model fitting revealed that the behavior of most of our subjects was best explained by the Influence model (**Fig. 3A**), while nearly one-third of the sample was best fitted by models with lower levels of strategic complexity (with fewer than 10% best-fitted by the standard Reinforcement Learning model). Moreover, subjects whose behavior was better predicted by the high SL model also exhibited higher values of their best-fitting Influence model parameter (λ) (B1: r = 0.7534, p = 6.757e-13; B2: r = 0.7535, p = 6.7276e-13; **Fig. 3B**). We conducted an extended computational analysis including additional models (see **Supplementary Material**, **MS2** and **MS3**). None of the tested variations of RL and belief-based models significantly

improved model fit (except for a second-order version of the Influence model, 2Inf, see **Fig. S6**), thus confirming that most subjects engaged in strategic learning, distributed along a gradient of strategic complexity (SL).

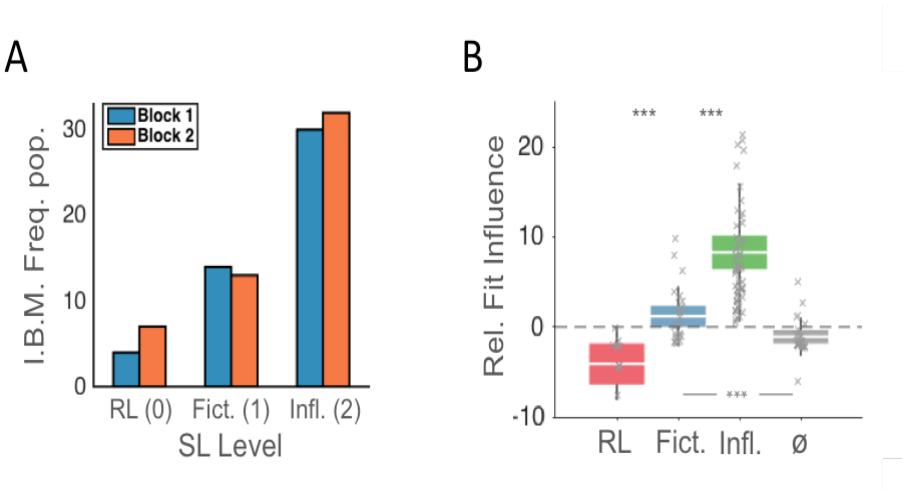

Figure 3. Strategic Learning heterogeneity captured by our computational approach (Experiment 2). The majority of participants engaged in Strategic Learning (SL > 0). (A) Individual Best Model (I.B.M.) frequency plot. Although the Influence model provided the best fit at the population level (not shown), at the individual level, most participants were best fit by high-SL models, while approximately one-third were best fit by models characterized by lower levels of strategic learning. (B) Population gradient of strategic learning sophistication. This plot displays the average relative fit quality of the Influence model (compared to the Reinforcement Learning model) for each SL group, based on their I.B.M. classification. Participants best fit by higher-level strategic learning models were increasingly better captured by the Influence model.

Strategic Learning, deviation from MSNE, and interaction dynamics. Next, we examined how SL levels influenced interaction dynamics. Player 2's SL level was negatively correlated with deviation from the Mixed Strategy Nash Equilibrium (MSNE) ($r = -0.6455$, $p = 8.48e\text{-}09$, using SL as a relative fit of the 2-Inf model). As suggested by our model-free analysis (**Fig. 4 A, B**), Players 2 SL level did not directly correlate with total points won (**Fig. 4C**), but instead with relative performance; Higher SL levels were associated with lower average relative losses (**Fig. 4A**). The higher Players 2 SL level, the closer their action distribution was to the MSNE (**Fig. 4B**). However, this was insufficient to overcome their structural disadvantage and increase absolute performance (**Fig. 4C**). In contrast to Players 2, Players 1 deviation from MSNE was not driven by their own SL level ($r = -0.0396$, $p = 0.7561$) but instead by their opponent's SL level ($r = 0.4940$, $p = 3.34e\text{-}05$, **Fig. 4B**). The higher Players 2 SL level, the worse Players 1 performed in both absolute ($r = -0.5826$, $p = 4.40e\text{-}07$) and relative terms ($r = -0.4650$, $p = 0.0001$). Higher-SL Players

2 deviated less from MSNE and reduced their reliance on the high-reward action, forcing Players 1 to adapt by increasing their own SL level. Given Players 1 structural advantage, the better they anticipated their opponent's behavior, the higher their relative earnings ($r = 0.4380$, $p = 0.0003$) and absolute performance. Similar results were obtained when running the correlation test analysis with the relative fit of the first-order Influence model. Using the Influence parameter (λ) as a measure of SL level or categorizing subjects into low and high SL groups (based on their best-fitting model) preserved the main statistical effects. These findings remain consistent also when comparing high- vs. low-SL groups (median split) within each role. To capture the simultaneous effects of both a subject's and their opponent's SL level on choice behavior, we conducted three GLM analyses. These confirmed that Players 2 behavior was primarily influenced by their level of strategic learning sophistication, whereas Players 1 behavior was mainly driven by the SL level of their opponent (**Table S1**).

Higher strategic learners are more unpredictable. This dynamic is further illustrated when examining how previous choices (own or the opponent's) affect current decisions. On average, subjects alternated choices every two trials, regardless of their role (**Fig. S7A**). However, only Players 2 tended to persist in selecting the high-reward action, taking less account of their opponent's previous choice (**Fig. S7 A, B**). The more Players 2 engaged in strategic learning, the more frequently they alternated their choices (**Fig. S7C**).

Correlation of SL with Additional Cognitive Tasks. We compared SL levels with individual performance in additional tasks and questionnaires. At the population level, only the CRT score (a widely used proxy for reasoning ability) was significantly higher for high-SL vs. low-SL participants (median split: Mann-Whitney U test = 313.5, $Z = 2.7678$, $p = 0.0056$; $r = 0.2572$, $p = 0.0402$), and only in Block 1. When comparing high vs. low SL subjects separately for each role in Block 1, we found that high-SL Players 1 had: (i) Higher CRT scores ($U = 54$, $Z = 2.8891$, $p = 0.0038$); (ii) Better performance in the Raven test ($U = 62$, $Z = 2.5389$, $p = 0.0111$); and (iii) Greater success in the Tower of London task (*t-test* = 2.0675, $p = 0.04918$; difficult condition: $U = 46$, $Z = 2.3271$, $p = 0.0199$). No significant correlation was found between cognitive tasks and SL level for Players 2 in either block. Additionally, no demographic differences (e.g., age, salary, education) or cognitive task performance differences were observed between players based on their assigned roles.

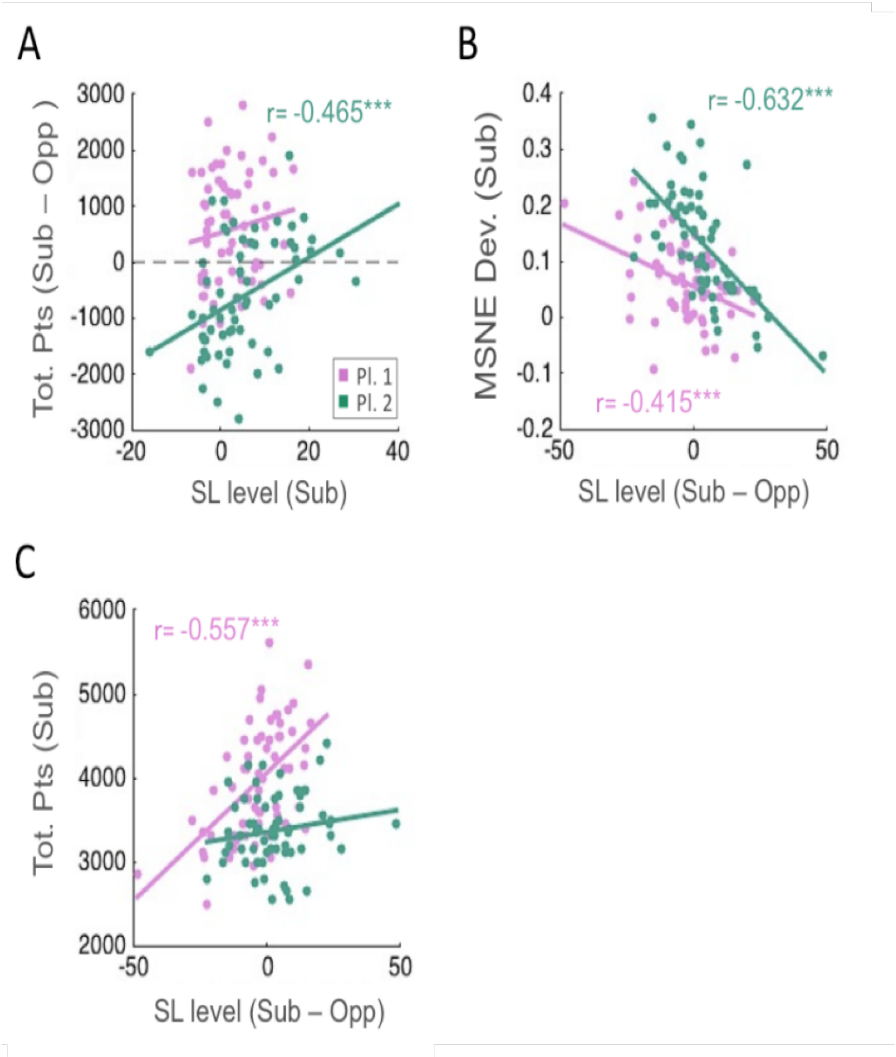

Figure 4. Model-based analysis (Experiment 2): The impact of Strategic Learning (SL) level on Players 2's ability to overcome their strategic disadvantage in the game. (A) The higher the SL level—measured here by the difference in fit between the second-order Influence model and the RL model, the more Players 2 reduced their disadvantage relative to their opponent. (B) The higher their SL level compared to their opponent, the closer they played to the MSNE. While both roles converged toward the equilibrium distribution, Players 2 deviated significantly more when they did not engage in strategic learning. (C) Although increasing their SL level helped Players 2 reduce their point deficit relative to their opponent, it did not significantly improve their overall performance. Their performance remained constrained by both the structure of the interaction and their own SL level.

SL and Strategic Reasoning (SR) in One-Shot Games. Our analysis (detailed in the Supplementary Material) shows no direct mapping between strategic learning (SL) and strategic reasoning (SR) at the population level. This finding suggests that distinct cognitive processes may underlie sophisticated strategic behavior in repeated games with feedback compared to static one-shot games without feedback.

Experiment 3: Human-machine interactions

To better characterize how participants adjust their behavior in response to their opponent (i.e., adaptive strategic learning), we conducted a behavioral study with humans in which we controlled the opponent's behavior. Participants played against a computer programmed to operate at a specific SL level throughout each block, facing a low-SL (Fictitious Play) opponent in one block and a high-SL (Influence) opponent in the other. On average, Players 1 scored more points (*t-test* = 8.5298, p = 2.7896e-14) and had a distribution of choices closer to the Mixed Strategy Nash Equilibrium (*t-test* = -4.5144, p = 1.322e-05, unequal variance) than Players 2. Our model-based analysis closely replicated both the distribution and the gradient of strategic learning (SL) levels observed among participants in Experiment 2 (**Fig. S8**). Additionally, as in Experiment 2, no difference in SL distribution was found between the two roles. For both roles, no significant difference was found in performance (total points or points difference with the opponent) between the two experiments. However, we observed a trend towards higher strategic learning when playing against algorithms. When comparing low vs. high SL (using a median split), Players 1 who engaged in strategic learning were found to have a higher SL level in human-machine (Experiment 3) vs. human-human (Experiment 2) interactions (relative fit 2-Inf: Mann-Whitney U test = 233, Z = 4.2082, p = 2.57e-05, λ parameter: U = 368, Z = 2.5495, p = 0.0108). Similar results were obtained when comparing the SL levels between subjects best fitted by the Influence model.

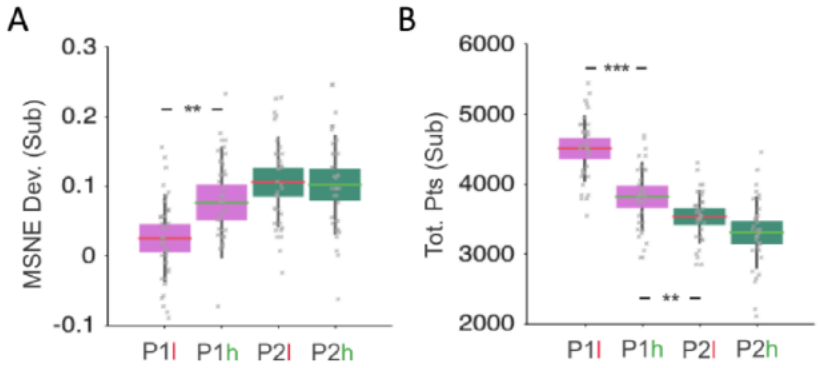

Figure 5. **Results for Experiment 3.** As hypothesized, only the subjects playing as Player 1 (P1) in the repeated game were affected in their choice behavior by the SL level of their (computerized) opponent ("l" for Low-SL, and "h" for High-SL of the computer opponent). (A) Players 1 frequency of choice is closer to the MSNE distribution when playing against the low-SL opponent compared to the high-SL opponent. No difference in the percentage of deviation from the MSNE distribution (p(a) = 1/3) was found between the two opponents for Players 2. (B) When playing against the low-SL opponent, Players 1 won, on average, more points in total than when playing against the high-SL opponent. No difference was found for Players 2.

As in Experiment 2, Players 2 were more consistently fitted by the same SL level models between the two opponents than Players 1 (P2 = 0.84, proportion of same low/high SL; P1 = 0.57; Fisher's exact test: p = 0.0259). Next, we tested our hypothesis regarding the specific effect of the opponent on participants' choice behavior, given the role they played in the experiment. As shown in **Fig. 5 (**and **Table S2)**, only Player 1 was affected by the SL level of their opponent, as predicted by the simulation analysis (**Fig. S9 A, B**).

DISCUSSION

This study integrates simulated agents, human participants, and human–machine interactions to examine how heterogeneous patterns of belief formation, specifically those associated with varying levels of strategic learning (SL) sophistication, interact with task structure to affect behavior and outcomes in repeated competitive games. Across three experiments, we demonstrate that structural asymmetries in the payoff matrix can produce systematic disparities in learning dynamics and performance. Our game design featured an embedded **strategic reverse-reward contingency** (SRRC), wherein one player's highest-payoff action (Player 2) was the one least prescribed by equilibrium play. This asymmetry introduces cognitive conflict: it demands that players suppress the natural tendency to pursue high-reward actions and instead adopt strategies informed by belief-based learning. Drawing from animal cognition paradigms [23,24], SRRC is conceptually analogous to reverse-reward tasks that require subjects to override reward-seeking impulses in favor of long-term gains. In our game, this tension challenges players to shift from reactive, reward-based learning to more deliberative, belief-based strategies. Our findings show that such transitions are critical for optimal behavior in dynamic settings. Consistent with this, players across roles exhibited varying degrees of SL sophistication, with a significant portion of behavior best captured by the **Influence model**, a computational framework incorporating recursive belief updating ("I think that you think that I think…"). This model, first introduced by Hampton et al. [5], captures how agents account for both the behavior and learning of their opponents.

Hampton et al. [5] found that brain activity in areas, such as the medial prefrontal cortex and posterior superior temporal sulcus, was more strongly associated with computations required by the Influence Model, supporting the idea that humans use higher-order strategic reasoning and mentalizing during social decision-making [9,24], going beyond what's captured by simpler models like Fictitious Play and RL.

In Experiment 1, simulations of agent-agent interactions showed that Player 1 consistently outperformed Player 2, despite both agents operating under identical SL algorithms. The advantage held across a range of SL levels, with Player 2 requiring disproportionately higher SL to match Player 1's performance. This asymmetry did not appear in structurally symmetric games like Matching Pennies, but was evident in the current paradigm and games like the Inspection Game [5] (see **Fig. S2**), indicating that performance differences arise not only from agent abilities but also from their strategic roles within the game.

Evidence for the Leader-follower dynamic. Experiment 2 extended these findings to human-human interactions. Once again, Player 1 outperformed Player 2, and computational modeling revealed widespread reliance on SL. Although the level of SL predicted relative improvement for Player 2, it did not eliminate their underlying disadvantage. A key emergent property of these interactions was a consistent **leader-follower dynamic**: Player 1 typically acted as an adaptive leader, strategically responding to and exploiting the predictable behavior of Player 2, who more often followed immediate reward signals. This asymmetry was not driven by differences in cognitive capability, as SL levels were evenly distributed across roles. Rather, it arose from the interaction between SL and the structural properties of the games, leading to differential outcomes. Similar findings are reported in Hampton et al. [5], where, in the Inspection Game, the employee holds a strategic advantage, exerting greater influence over the employer. This asymmetry is reflected in increased activity in the medial prefrontal cortex (mPFC) for employees compared to employers during switching versus non-switching trials [5]. Our study provides a theoretical and computational account of this phenomenon, which is not explored in [5].

In Experiment 3, we examined human-machine interactions by manipulating the strategic learning (SL) level of artificial agents. Our findings show that human Player 1 participants adapted their strategies to the sophistication of their computerized opponents, demonstrating sensitivity to the

agents' SL level. In contrast, human Player 2 participants exhibited limited adaptation, remaining structurally constrained in their ability to fully exploit strategic opportunities. These results contribute to the literature on human-machine interaction in repeated games with mixed-strategy equilibria [25-27], demonstrating how structural features of the game, not just cognitive limits, can restrict adaptive behavior.

Our computational modeling provided strong support for the Influence model across all experimental conditions. To further validate this finding, we compared the Influence model with two alternative models: Experience Weighted Attraction (EWA) [28,29] and two-period fictitious play (FP2) [30]. While both alternatives offer richer representations than classical fictitious play, by integrating counterfactual reasoning and pattern detection, respectively, neither model outperformed the Influence model in capturing strategic sophistication. This suggests that the recursive, belief-based structure of the Influence model reflects a qualitatively distinct cognitive mechanism, one that underlies higher-order learning and mentalizing in social decision-making.
Importantly, our modeling also clarified the differential effects of SL across roles. For Player 2, higher SL was associated with reduced deviation from MSNE and smaller losses, but not with performance gains, indicating a ceiling effect in their ability to benefit from strategic learning. For Player 1, SL had minimal impact on equilibrium alignment but moderated the extent to which they could capitalize on their opponent's behavior, suggesting a more adaptive and opportunistic use of SL.

Cognitive skills and strategic learning in games. Interestingly, cognitive assessments showed that higher SL in Player 1 correlated with better performance on tasks measuring cognitive flexibility and reasoning (e.g., CRT, Raven's matrices), but these associations were absent in Player 2. This implies that the structural burden imposed on Player 2 may blunt the expression or impact of cognitive abilities. Furthermore, SL was distinct from traditional measures of strategic reasoning in one-shot games.

Insights for behavioral game theory. Our findings also contribute to behavioral game theory by extending the "own payoff effect" [31] to a nominally symmetric game design. In our study, deviations from MSNE were consistent with an overvaluation of high-reward actions, particularly by Players 2, and mirrored earlier findings showing how payoff salience can distort behavior away

from equilibrium strategies. Interestingly, increasing the highest payoffs in our game matrix produces opposite effects (specifically, greater deviations from equilibrium, **Fig. S5**) than those predicted by the Quantal Response Equilibrium (QRE) [32]. This suggests that the determinants of out-of-equilibrium behavior in our setting differ fundamentally from those assumed by QRE.

Our work builds on seminal research on learning in games [33] by demonstrating the **relevance of adaptive learning models** [34] in explaining behavior in repeated interactions. In contrast to Erev and Roth [33], our findings reveal a systematic advantage of higher-order strategic processes (recursive belief updating).

We contribute to the behavioral literature on strategic interaction, examining whether humans adhere to minimax or mixed-strategy equilibria in zero-sum and constant-sum games played in the laboratory and real-life settings (35-43). Our findings show that conflicts between high-reward actions and their equilibrium probabilities are key sources of deviation from mixed strategies. This suggests that basic reward-learning mechanisms, often overlooked in behavioral game theory, play a central role in driving out-of-equilibrium behavior in strategic interactions.

Algorithmic fairness in competitive "synthetic" environments. The implications for artificial intelligence and algorithmic fairness are significant. Simulated agent interactions revealed that even when agents possess identical computational resources, structural features of the game can generate persistent inequalities. Player 2 agents required more complex strategies to achieve comparable outcomes, raising concerns about fairness in AI evaluation and deployment. These findings suggest that the success or failure of AI agents may hinge not only on algorithmic sophistication but also on the roles or constraints imposed by the environment, echoing broader concerns about **context-sensitive bias in AI systems** [44,45].

Human-machine interactions in competitive environments. From a human–machine interaction perspective, our results reveal the adaptive capacity of human learners when interacting with artificial agents, as well as the limits imposed by structural asymmetries. Even when humans detect and adjust to machine sophistication, their performance may remain bounded by the role they occupy. This insight is crucial for **designing equitable human-AI systems** in competitive contexts, where asymmetries in information, role, or feedback can reinforce performance

disparities. Ensuring fairness in such settings requires attention not only to algorithmic parity but also to structural game design and interactional dynamics.

Conclusion. Our study offers a mechanistic account of how individuals engage in strategic belief updating and learning in repeated competitive interactions. Through a combination of behavioral experiments and computational modeling, we highlight the interplay between strategic learning sophistication, game structure, and adaptive behavior, providing insights into the cognitive and contextual determinants of social decision-making. These insights advance both theoretical understanding and practical applications in behavioral game theory, human-AI systems, and the design of fair algorithmic environments.

METHODS

Experimental Task: The Game. The stage game is a two-by-two (two players, two actions) normal-form game with a unique Mixed Strategy Nash Equilibrium (MSNE, **Fig. 1A**). The MSNE is a theoretical solution in which each player adopts a probability distribution over available actions such that no player has an incentive to unilaterally deviate, provided all others adhere to their strategies. In this equilibrium, each player's strategy is calibrated to make the opponent indifferent among their available actions, thereby justifying randomization. The MSNE prescribes that Player 1 should play action *a* with a probability of 1/3 and action *b* with a probability of 2/3, while Player 2 should play action *A* with a probability of 2/3 and action *B* with a probability of 1/3. The expected payoffs at the MSNE are the same for both players. However, the highest payoff (100 points) corresponds to the action that Player 1 should theoretically play most frequently and the action that Player 2 should play least frequently. This feature introduces a strategic asymmetry between the two roles.

Experiment 1: Machine Against Machine – Model Simulation and Prediction. To predict how the strategic asymmetry of our payoff matrix affects behavior, we simulated computerized agents playing a repeated version of our game in each of the two roles, using different levels of strategic sophistication. This simulation analysis allows us to assess the robustness of our design and generate precise predictions regarding how individual differences in strategic learning influence

behavior in our experimental setting. To account for inter-individual variation in strategic learning (SL), we adopted three computational models with varying levels of strategic sophistication (**Model Space 1, MS1** in Supplementary Material). **Q-Learning (no-SL):** A simple reinforcement learning algorithm that updates choices solely based on past outcomes [33]. In social settings, RL models assume that agents do not form beliefs about other players' behavior. **Fictitious Play (low SL):** Fictitious Play (low SL): This model computes a best response to the probability of each opponent's choice, determined by their historical actions [46, 47]. Essentially, it assumes that an opponent's future behavior can be estimated based on the distribution of their past choices. **Influence Model (high SL):** A second-order fictitious play model that considers the influence of a player's past choices when estimating the opponent's probability of play [5]. The Influence Model builds upon the fictitious play framework by incorporating a second layer of analysis in belief formation. It not only considers the opponent's history but also factors in how the player's own past choices may have influenced the opponent's behavior. In the Influence model, the opponent's learning rate is captured by the influence parameter λ (ranging from 0, indicating equivalence to Fictitious Play, to 1, representing a full Influence Model), which quantifies the degree to which the agent's actions affect the opponent's behavior. Meanwhile, the agent's learning rate is represented by η. We expanded the initial model space by incorporating a **hybrid model** characterized by two main parameters: a λ parameter that captures the degree to which the agent believes its own past choices influence the opponent's behavior, and a κ parameter that reflects the agent's propensity to engage in second-order belief updating (i.e., Influence) relative to first-order belief updating (i.e., fictitious play). While conceptually close to the Influence model that varies the λ parameter, the hybrid model introduces an added layer of control that strengthens the validity of our simulation analysis. For more details, see the Computational Model Section in the Supplementary Material. Each simulation consisted of two computerized agents, each assigned one of the two roles and modeled by one of the three learning algorithms, playing against each other for 100 repetitions of the stage game. Additionally, we ran similar simulations for two different games: a matching pennies game and the inspection game from Hampton et al. [5] (see **Fig S2**). The computational analysis of agent-agent interactions across the three games offers a detailed characterization of strategic learning dynamics under varying competitive conditions.

Experiment 2: Human Against Human

Participants. Sixty-four participants (29 males, 35 females; mean age: 27.1 ± 9.4 years) took part in the experiment. They were students at the University of Lyon 1, France, who had voluntarily joined the recruitment system. All participants provided written informed consent, and the study was approved by the French National Ethical Committee. Participants were right-handed, medication-free, had normal eyesight, and had no history of neurological disorders.

Experimental Design. This experiment involved repeated interactions between human participants. At the start, each participant was randomly assigned one of the two roles. Each subject interacted with two different human opponents, one in each of two trial blocks, each consisting of 100 repetitions of the stage game with full choice feedback (**Fig. S1 A, B**). The opponents were randomly selected from participants assigned the opposite role. Points earned in each trial were accumulated across the block and converted to final earnings, with 1,000 points corresponding to 1€, in addition to a 5€ show-up fee. Participants were initially instructed about the two stimuli representing their available actions, the game's payoff structure, and the stimulus-outcome contingencies in the payoff matrix. Each action was represented by a different colored circle, randomly chosen from four possible colors (all controlled for luminance). These colors were randomly assigned to each pair of subjects in the first block and remained unchanged in the second block, thereby constraining the re-matching process. During each trial, both players had three seconds to select one of the two colors displayed on the left and right of the screen (in randomized order across trials). The chosen color was highlighted for one second as choice feedback. Four seconds after trial onset, both players were simultaneously shown the outcome feedback for their own choice and that of their opponent, displayed for three seconds. The outcome feedback screen presented the payoff matrix, with the following features: i) The matrix was flipped depending on the player's role to ensure they were always presented as the row player. ii) The cell corresponding to the players' chosen actions was highlighted. iii) The points won by the subject were displayed in turquoise (**Fig. S1B**). This display minimized framing effects while ensuring that participants remained aware of the underlying payoff structure of the game. We also asked participants to complete an additional task consisting of four different types of 2×2 static (one-shot) games [48]. This was designed to test the hypothesis developed in [49] that participants with a high SL level in the repeated game, captured using our computational approach, would also exhibit higher strategic reasoning in non-repeated (static) games without feedback. Strategic reasoning (SR) was

measured by their ability to conform to equilibrium play in games that were not repeated and provided no feedback (see Supplementary Material). All subjects returned to the lab one week later to complete a series of cognitive tasks. Details of this follow-up experiment and the specific tasks administered (e.g., Cognitive Reflection Test, CRT, Raven's Progressive Matrices, Tower of London, and n-back working memory task) are provided in the Supplementary Material. These tasks were used to assess participants' general cognitive abilities and the specific tendency to inhibit intuitive responses to engage in sophisticated reasoning (CRT).

Computational Modeling. To assess the participants' level of strategic learning, we applied the computational approach introduced in the simulation analysis (Experiment 1, Model Space 1, MS1). Specifically, we fitted each participant's choices in both trial blocks to three computational models, representing increasing levels of strategic sophistication: i. Q-Learning (no-SL), learning is based solely on past outcomes. ii. Fictitious Play (low SL), responds to the probability of an opponent's actions based on historical data. iii. Influence Model (high SL), incorporates the influence of one's past choices on the opponent's future decisions [5]. The underlying assumption was that the best-fitting model for each participant would reflect their level of strategic engagement: the more complex the best-fitting model, the higher their strategic learning level. As detailed in the Supplementary Information (**Model Spaces 2** and **3**), we also tested additional models to validate the robustness of our computational approach. For the computational analysis in Experiment 2, we expanded the initial model space from Experiment 1 by incorporating additional reinforcement learning and belief-based learning models. These included Generalized Reinforcement Learning, Counterfactual Reinforcement Learning, Weighted Fictitious Play, the Influence Extension model (2-Influence) (Model Space 2, MS2), the Experience Weighted Attraction Model (EWA [23,24]), and a patterned variation of Fictitious Play (FP-2; [25]) (details in **Supplementary Material**).

Experiment 3: Human Against Machine

Participants. Seventy-six participants (36 males, 40 females; ages 18–30) took part in this experiment. They were students at the University of Trento, Italy, who had voluntarily joined the Cognitive and Experimental Economics Laboratory (CEEL) recruitment system. All participants were right-handed, medication-free, had normal eyesight, and had no history of neurological

disorders. The experiment was approved by the Ethics Commission of the University of Trento, and informed consent was obtained from each subject before participation. Data collection was conducted blind to experimental conditions.

Experimental Design. The experimental design remained unchanged: participants were randomly assigned to one of the two roles in the same game (**Fig. 1A**), following the same trial structure and timing as in the previous experiment. Each participant played two blocks of 100 trials, but instead of playing against human opponents, they played against two computerized learning agents: Fictitious Play (low SL) and Influence Model (high SL). Each participant interacted with both agents, with the order of interaction randomized across participants, and was informed that the opponent was a computer algorithm designed to respond adaptively to their choices. To ensure consistency with the previous experiment, we set the parameters of the two computerized opponents to match the average best-fitting values obtained from human participants in Experiment 2 (details in **Supplementary Material**).

REFERENCES

1. Lee D, Seo H (2016) Neural basis of strategic decision making. *Trends Neurosci* 39:40–48.
2. Amodio DM, Frith CD (2016) Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 7:268–277.
3. McCabe K, Houser D, Ryan L, Smith VL, Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange. *Proc Natl Acad Sci USA* 98:11832-11835.
4. Yoshida W, Seymour B, Friston KJ, Dolan RJ (2010) Neural mechanisms of belief inference during cooperative games. *J Neurosci* 30:10744–10751.
5. Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci USA* 105:6741–6746.
6. Camerer C (2003) Behavioral game theory: Experiments in strategic interaction. (Princeton University Press, Princeton, NJ).
7. Nagel R (1995) Unraveling in guessing games: an experimental study. *Am Econ Rev* 85:1313–1326.

8. Bhatt M, Camerer CF (2005) Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games Econ Behav* 52:424–459.
9. Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc Natl Acad Sci USA* 106:9163–9168.
10. Camerer CF, Ho TH, Chong JK (2004) A cognitive hierarchy model of games. Q J Econ 119:861–898.
11. Stahl DO, Wilson PW (1995) On players' models of other players: Theory and experimental evidence. *Games Econ Behav* 10:218–254.
12. Costa-Gomes M, Crawford V, Broseta B (2001) Cognition behavior in normal-form games: an experimental study. *Econometrica* 69:1193–1235.
13. Fudenberg D, Levine DK (1998) The theory of learning in games. (MIT Press, Cambridge, MA).
14. Camerer CF, Ho TH, Chong JK (2004) Behavioral game theory: Thinking, learning and teaching. In Advances in Understanding Strategic Behaviour (Palgrave Macmillan, London).
15. Sutton RS, Barto AG (1998) Introduction to Reinforcement Learning. (MIT Press, Cambridge, MA).
16. Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol* 22:1075–1081.
17. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND (2015) Model-based choices involve prospective neural activity. *Nat Neurosci* 18:767–772.
18. Ruff CC, Fehr E (2014) The neurobiology of rewards and values in social decision making. *Nat Rev Neurosci* 15:549–562.
19. Devaine M, Hollard G, Daunizeau J (2014) The social Bayesian brain: does mentalizing make a difference when we learn? *PLoS Comput Biol* 10:e1003992.
20. Hill CA, Suzuki S, Polania R, Moisa M, O'Doherty JP, Ruff CC (2017) A causal account of the brain network computations underlying strategic social behavior. *Nat Neurosci* 20:1142–1149.
21. Konovalov A, Hill CA, Daunizeau J, Ruff CC (2021) Dissecting functional contributions of the social brain to strategic behavior. *Neuron* 109:3323–3337.
22. Boysen, S.T., and Berntson, G.G. (1995). Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (Pan troglodytes). J. Exp. Psychol. *Anim. Behav. Process.* 21:82–86.

23. Beran, M.J. (2023). I'll (not) take that: The reverse-reward contingency task as a test of self-control and inhibition. *Learn. Behav.* 51: 9–14.
24. Nagel, R., Brovelli, A., Heinemann, F., & Coricelli, G. (2018). Neural mechanisms mediating degrees of strategic uncertainty. *Soc. Cogn. Affect. Neurosci.* 13: 52–62.
25. Messik DM (1967) Interdependent Decision Strategies in Zero-Sum Games: A Computer-Controlled Study. *Behav Sci* 12: 33-48.
26. Fox J (1972) The learning of strategies in a simple, two-person zero-sum game without saddlepoint. *Behav Sci* 17:300–308.
27. Shachat J, Swarthout TJ (2004) Do we detect and exploit mixed strategy play by opponents? *Math Methods Oper Res* 59:359–373.
28. Camerer CF, Ho TH (1999) Experience-weighted attraction learning in normal form games. *Econometrica* 67:827–874.
29. Camerer CF, Ho TH, Chong JK (2002) Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *J Econ Theory* 104:137–188.
30. Spiliopoulos L (2012) Pattern recognition and subjective belief learning in a repeated constant-sum game. *Games Econ Behav* 75:921–935.
31. Goeree JK, Holt CA (2001) Ten little treasures of game theory and ten intuitive contradictions. *Am Econ Rev* 91:1402–1422.
32. McKelvey RD, Palfrey TR (1995) Quantal response equilibria in normal form games. *Games Econ Behav* 10:6–38.
33. Erev I, Roth AE (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am Econ Rev* 88:848–881.
34. Selten, R. (1991). Evolution, learning, and economic behavior. Games and Economic Behavior, 3: 3–24.
35. Lieberman B (1960) Human behavior in a strictly determined 3 × 3 matrix game. *Behav Sci* 5:317–322.
36. Brayer AR (1964) An experimental analysis of some variables of minimax theory. *Behav Sci* 9:33–44.
37. O'Neill B (1987) A non-metric test of the minimax theory of two-person zero-sum games. *Proc Natl Acad Sci USA* 84:2106–2109.

38. Brown JN, Rosenthal RW (1990) Testing the minimax hypothesis: A re-examination of O'Neill's game experiment. *Econometrica* 58:1065–1081.
39. Budescu DV, Rapoport A (1992) Generation of random series in two-person strictly competitive games. *J Exp Psychol Gen* 121:352–363.
40. Walker M, Wooders J (2001) Minimax play at Wimbledon. *Am Econ Rev* 91:1521–1538.
41. Palacios-Huerta I (2003) Professionals play minimax. *Rev Econ Stud* 70:395–415.
42. Palacios-Huerta I, Volij O (2008) *Experientia docet*: Professionals play minimax in laboratory experiments. *Econometrica* 76:71–115.
43. Levitt SD, List JA, Reiley DH (2010) What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments. *Econometrica* 78:1413–1434.
44. Mullainathan, S., & Obermeyer, Z. (2017). Does machine learning automate moral hazard and error? *Am. Econ. Rev.* 107: 476–480.
45. Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *Am. Econ. Rev.* 110: 3267–3297.
46. Brown GW (1951) Iterative solutions of games by fictitious play. In Koopmans TC, ed. Activity Analysis of Production and Allocation (Wiley, New York), pp 374–376.
47. Robinson J (1951) An iterative method of solving a game. *Ann Math* 54:296–301.
48. Polonio L, Di Guida S, Coricelli G (2015) Strategic sophistication and attention in games: An eye-tracking study. *Games Econ Behav* 94:80–96.
49. Griessinger T, Coricelli G (2015) The neuroeconomics of strategic interaction. *Curr Opin Behav Sci* 3:73–79.