BELIEFS: A Hierarchical Theory of Mind Model based on Strategy Inference

Giorgio L. Manenti (g.manenti@uke.de)
Sepideh Khoneiveh (s.khoneiveh@uke.de)
Oleg Solopchuk (o.solopchuk@uke.de)
Jan Gläscher (glaescher@uke.de)

Department of Systems Neuroscience, University Medical Center Eppendorf Hamburg
20246 Hamburg, Germany

Abstract

Theory of Mind (ToM) refers to the capacity to infer others' latent mental states, such as intentions, beliefs, and strategies, and use these inferences to predict behavior. A defining characteristic of ToM is its recursive nature: individuals reason not only about what others are thinking, but also about what others think about them. Most computational models of ToM adopt a hierarchical structure in which Level-0 (L0) agents are assumed to follow simple, fixed heuristics (e.g., Win-Stay-Lose-Shift, WSLS) without mentalizing. However, this assumption overlooks the diversity of non-mentalizing strategies exhibited in human behavior, such as imitation or tit-for-tat, which do not conform to WSLS yet require no recursive reasoning. To address this limitation, we introduce a novel ToM framework (*BELIEFS*) that flexibly infers latent L0 strategies from behavior rather than relying on predefined heuristics. We evaluated the model in four classic dyadic games: Matching Pennies, Prisoner's Dilemma, Bach or Stravinsky, and Stag Hunt, manipulating model's learning rates and the volatility of L0 strategy switching. Predictive accuracy was assessed using cumulative negative log-likelihood (NLL) of opponent's next choice and compared against both a ToM model that assumes only WSLS at L0 and chance-level performance. Our model outperformed both baselines, particularly under low-volatility conditions and at intermediate learning rate. Moreover, to evaluate strategy inference, we computed trial-wise confusion matrices and Cohen's κ between inferred and true L0 strategies, reaching significantly above-chance classification. We further tested the model's ability to distinguish between action sequences generated by the opponent's true Theory of Mind (ToM) level (L0 vs. L1) and those generated using an incorrect ToM level. The model assigned lower negative log-likelihoods (NLLs) to sequences from the true level, suggesting an indirect method for identifying the opponent's actual ToM level. Finally, we assessed whether the model effectively tracks behaviorally distinguishable action probabilities across ToM levels. Using Fisher-transformed correlations between model-generated action probabilities at L0, L1, and L2, we found significant dissimilarities, especially in competitive games. In summary, our model introduces a flexible, probabilistic approach to Theory of Mind that captures both surface-level strategy use and recursive reasoning depth. By jointly tracking dynamic beliefs over L0 strategies and ToM levels, the model adapts to behavioral shifts and outperforms static heuristics. These advances provide a powerful framework for modeling human behavior in interactive contexts, with implications for both human-human and human-machine interaction research.

Introduction

Theory of Mind (ToM) refers to the capacity to infer the unobservable mental states of others (Langley, Cirstea, Cuzzolin, & Sahakian, 2022; Premack & Woodruff, 1978; Wimmer & Perner, 1983), such as their intentions, beliefs, desires, and goals, and to use this information to predict behavior (Cho, Kamkar, & Hosseini-Kamkar, 2022; Schneider, Slaughter, & Dux, 2017). A central feature of ToM is its recursive nature (C. F. Camerer, Ho, & Chong, 2004; Stahl, 1993): individuals can represent not only what others are thinking, but also what others think about themselves (e.g.,

“I believe that you believe I will do X”). This recursive reasoning is typically organized into hierarchical levels (C. Camerer, Ho, & Chong, 2003; Zhen & Yu, 2021). At Level-0, an agent acts without considering the mental states of others. At higher levels, agents model others as reasoning one level below themselves, forming increasingly complex mental simulations.

ToM is foundational to human social interaction (Estes & Bartsch, 2017) and is increasingly important in human–machine contexts, such as collaborative AI (Langley et al., 2022), autonomous agents (Montes, Luck, Osman, Rodrigues, & Sierra, 2023; Rabinowitz et al., 2018), and social robotics (Esterwood & Robert, 2023; Krach et al., 2008). At a general level, ToM enables agents to construct mental models of others that support accurate predictions of their future behavior. These models range from non-mentalizing strategies (Level-0) to more sophisticated recursive reasoning (Level-1 and above), where an agent simulates the thought processes of others who, in turn, may be simulating them.

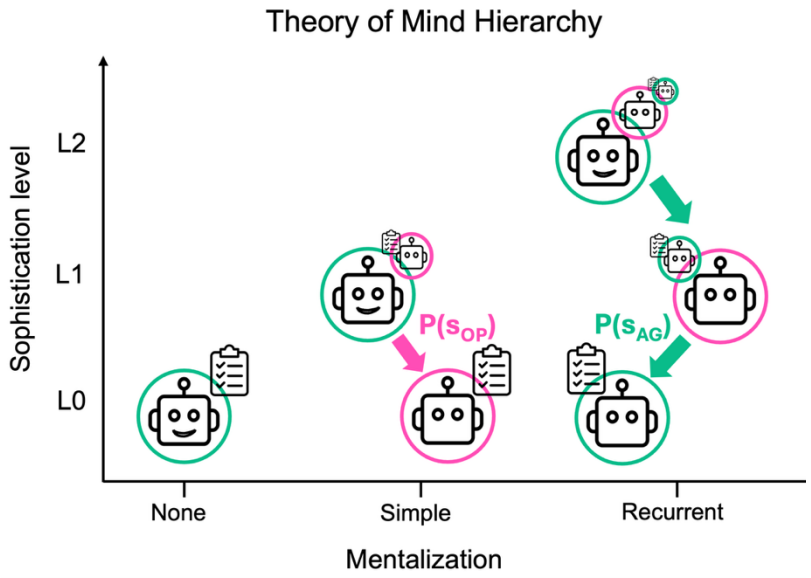

Figure 1: Hierarchy over ToM levels. At Level 0, the agent (AG) follows action strategies without considering the strategies used by the opponent (OP). At Level 1, the agent estimates a probability distribution $P(s_{OP})$ over opponents actions, marginalizing over the belief over strategies. At Level 2, the agent simulates the opponent’s best response, assuming the opponent forms a belief about the agent’s strategies $P(s_{AG})$, and then best responds to that. All mentalizing occurs from the agent’s perspective.

Several computational models have been developed to formalize the hierarchical complexity of mentalizing (Devaine, Hollard, & Daunizeau, 2014; Yoshida, Dolan, & Friston, 2008). These models typically begin with a Level-0 (L0) agent, who knows the rules of the game and selects actions based solely on environmental feedback, without attributing intentionality to the opponent. In this framework, the opponent’s behavior is treated as part of the environment, not as a product of strategic reasoning.

Most models assume that L0 agents employ simple reinforcement-based heuristics, such as the Win-Stay, Lose-Shift (WSLS) strategy. This strategy predicts that an agent will repeat a choice following a reward and switch following a loss. While convenient computationally, this assumption underestimates the variability of human decision-making at even the most basic levels. For example, in repeated economic games such as the Iterated Prisoner’s Dilemma, Tit-for-Tat, a reactive strategy that mimics the opponent’s previous move, has been shown to be highly effective, even though it does not require recursive reasoning or Theory of Mind (Axelrod, 1980). Crucially, this strategy also

treats the opponent as an environmental regularity but does not conform to the WSLS pattern. These observations highlight a critical limitation in many existing models: they rely on a narrow and static conception of Level-0 behavior, which fails to capture the richness of actual human strategies.

Our model addresses this limitation by allowing for flexible L0 strategy inference. Rather than hardcoding a specific heuristic, the model infers the opponent's L0 strategy from observed behavior and uses this inference as the foundation for building higher levels of reasoning. Specifically, in our framework, a Level-1 (L1) agent models the opponent as an L0 agent. She attempts to infer which L0 strategy the opponent is using and then computes the best response to that inferred behavior. Importantly, the best response is done on the integration of the beliefs about the L0 strategy. At Level-2 (L2), the agent models the opponent as a Level-1 reasoner who, in turn, models the agent as L0. Thus, the L2 agent simulates how her opponent is reasoning about her own behavior and selects a best response to that recursive inference. This iterative structure continues at higher levels, where an agent at Level-k simulates an opponent at Level-k–1, forming a recursive chain of "best responses to best responses."

This hierarchical organization carries several structural implications for how recursive reasoning unfolds in the model. First, the agent's mental model of the *opponent is always situated one level below* her own, as each level of reasoning is constructed by computing a best response to the simulated behavior at the lower level. Second, these *recursive simulations are built iteratively*: the agent begins with an assumed Level-0 representation and constructs higher levels step by step by nesting best responses to the previous level's behavior. Third, and crucial point, the *attribution of Level-0 strategies alternates across levels in a systematic way*. At odd-numbered levels (e.g., Level-1, Level-3), the agent models the hierarchy starting from opponent's Level-0 strategies, whereas at even-numbered levels (e.g., Level-2, Level-4), the agent considers her own behavior as the base of the levels reasoning (see Figure 1). Finally, because different ToM levels entail different simulated beliefs and best responses, the *model generates distinct predictions of behavior at each level*. These differing predictions can be compared to observed choices, thereby enabling inferences about the likely ToM level at which each player is operating.

Critically, this structure makes the accurate detection of the L0 strategy central to the overall predictive power of the model. All higher-order reasoning depends on this base-level inference. This is the key departure point of our model from earlier recursive ToM frameworks (Devaine et al., 2014; Yoshida et al., 2008), which typically fix the L0 behavior to WSLS or its variation to simple reinforcement learning policy. These static assumptions ignore the empirical fact that real human strategies at L0 can vary substantially and may evolve over time.

Building on this recursive architecture, our framework generalizes traditional ToM models by replacing fixed heuristics at Level-0 with flexible, probabilistic inference over multiple candidate strategies. These beliefs are continuously updated and integrated to guide behavior prediction. Each ToM level then generates a best response to the aggregated behavior predicted at the level below. In parallel, the model infers the opponent's depth of reasoning by maintaining and updating a belief distribution over ToM levels. This joint inference over strategies and ToM levels, grounded in Bayesian-style belief updating, is the foundation of our model, which we term **B**ayesian **E**stimation of **L**evels and **I**ntegration of **E**vidence over **F**lexible **S**trategies (**BELIEFS**).

Building on these L0 predictions, the model computes higher-level choices as best responses to simulated opponent behavior, with each Theory of Mind (ToM) level recursively modeling and responding to the level beneath it. While inspired by earlier ToM models developed for matrix games

(De Weerd, Diepgrond, & Verbrugge, 2018; de Weerd, Verbrugge, & Verheij, 2013), the current framework refines and generalizes the belief integration mechanism, addressing inconsistencies in previous approaches and enabling more robust and adaptive mentalizing.
A detailed example of how beliefs are computed by a Level-2 agent, integrating both L0 strategy beliefs and ToM-level reasoning, is illustrated in Figure 2.

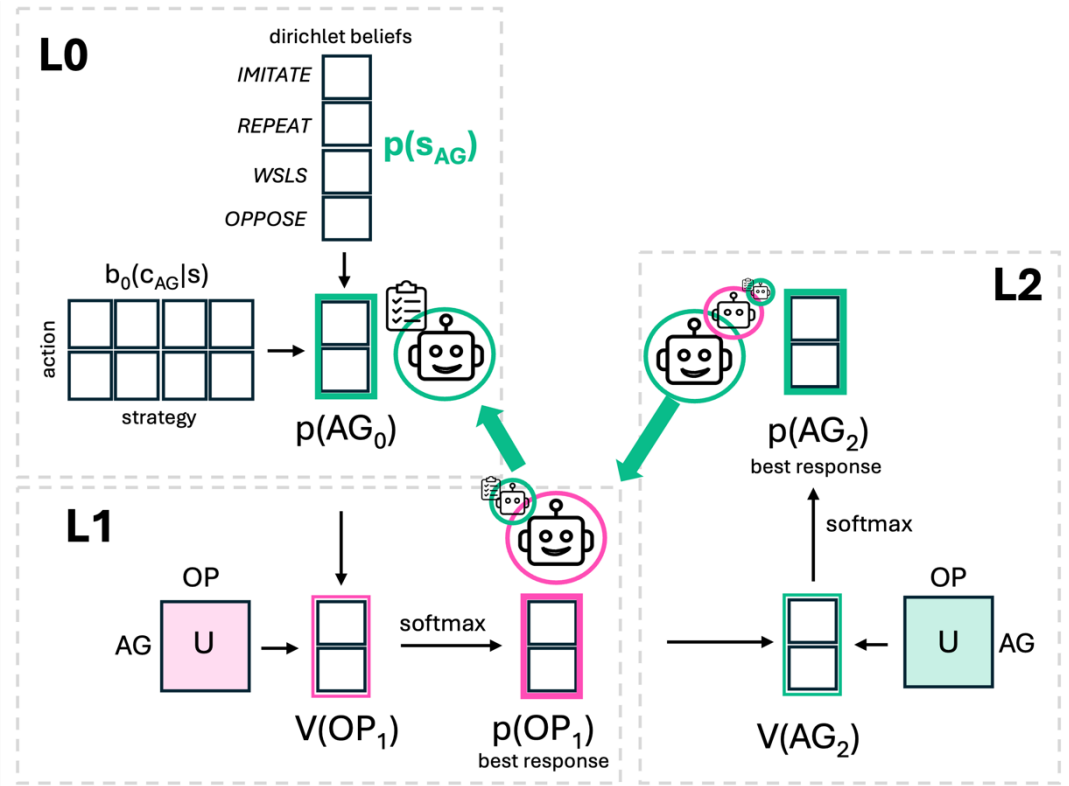

Figure 2 – **BELIEFS** in action: hierarchical belief inference and action prediction from the perspective of a Level-2 (L2) agent. This figure illustrates the full recursive structure of the **BELIEFS** model as implemented by an L2 agent engaged in a two-choice game. To compute her action probabilities, the L2 agent simulates the reasoning process of her opponent, assumed to be a Level-1 (L1) agent. The simulation begins by projecting down to the Level-0 (L0) layer, where the L1 opponent models the agent as an L0 player and infers which strategy *s* she is using. This inference is represented by a belief distribution over strategies $p(s)$, computed via the Dirichlet-based evidence accumulation process (see Methods). Using these beliefs and the strategies contingencies $b_0(C_{AG}|s)$, the L1 opponent integrates across strategies to derive a predicted action probability for the L0 agent, denoted $p(AG_0)$. The L1 opponent then computes expected values for her own possible actions $V(OP_1)$ by weighting her utility function U with $p(AG_0)$. These values are transformed into a probabilistic choice distribution $p(OP_1)$ using a softmax decision rule. The L2 agent, in turn, simulates this entire process to estimate $p(OP_1)$, and best-responds by calculating her own expected values using her utility function and computing the resulting choice probabilities $p(AG_2)$. Crucially, all components of this hierarchical simulation—beliefs, values, and actions—are internal constructions of the L2 agent.This schematic exemplifies how the BELIEFS model recursively constructs Theory of Mind reasoning by integrating beliefs over both strategy space and ToM levels, allowing for flexible and adaptive prediction of behavior.

Results

Predicting L0 Opponent's action and Tracking L0 strategies

A central aim of Theory of Mind (ToM) models is to predict others' behavior by inferring their latent mental states, such as beliefs, intentions, and strategies, especially in interactive decision-making contexts. In our study, we evaluate the predictive capabilities of our BELIEFS model by assessing its performance in forecasting an opponent's next choice under varying levels of environmental

volatility. Volatility here refers to the probability that the opponent changes their underlying decision-making strategy (specifically, their Level-0 strategy) on a given trial. We examine two theoretically informative extremes: Low volatility (10% chance of strategy change per trial) and High volatility (40% chance of change per trial). These extremes are chosen in order to approximate plausible human behavioral variability (Traulsen, Semmann, Sommerfeld, Krambeck, & Milinski, 2010).
To systematically evaluate model performance, we simulate four classic dyadic games over 40 trials per pair: Matching Pennies, Prisoner's Dilemma (PD), Bach or Stravinsky (BoS), and Stag Hunt (SH). Each game sequence is simulated across 100 independent dyads. Within each simulation, we manipulate the model's strategy learning rate parameter across four levels (0.1, 0.3, 0.5, 0.9) to explore how adaptive updating influences predictive accuracy under different volatility conditions.

We assessed the predictive performance of our model by computing the cumulative negative log-likelihood (NLL) of the opponent's choices across. As a benchmark, we compared our model against both a chance-level baseline and a model that assumes that the opponent follows classical Win-Stay-Lose-Shift at L0. In the WSLS implementation, predicted choices followed a deterministic win/loss rule, passed through a softmax function with a fixed temperature of 1 to ensure consistency with our model's parametrization.

Across all four game types and under both low and high volatility, our model consistently outperformed the WSLS heuristic in predicting opponent choices, especially at intermediate learning rates. For instance, in the Matching Pennies (MP) game under low volatility, the model achieved significantly lower NLLs than WSLS at a learning rate of 0.9 (mean difference = 4.52, $t(99) = 2.41$, $p = 0.02$, Hedges' $g = 0.43$). A similar pattern held under high volatility, with the exception of learning rate 0.9, where the model underperformed relative to WSLS. To investigate this discrepancy at high learning rate, we analyzed trial-level contributions to NLL. We found that the model rapidly reaches high confidence about opponent's strategy, which consequently leads to pay high penalty at strategy-switch trials, causing mispredictions about opponents next choice (Supplementary Figure S1). This suggests that under highly volatile environment, high learning rates may cause overreactive belief updates, diminishing the benefits of structured inference. All reported p-values were corrected for multiple comparisons using the FDR procedure (Benjamini & Hochberg, 1995).

The model significantly outperformed chance-level prediction (log(0.5)) across most learning rates and volatility conditions (Figure 3). For instance, at a learning rate of 0.1 under low volatility, the model's average NLL was well below chance (mean NLL = 25.33, $t(99) = 5.34$, $p < 0.001$, Hedges' $g = 0.53$). Similar results were observed at learning rates 0.3 and 0.5. At a learning rate of 0.9 under low volatility, the model remained above chance for PD, SH, and BoS, but not for MP ($p = 0.10$). Under high volatility, performance fell below chance across all games ($p > 0.05$), suggesting that aggressive belief updating hinders accurate strategy inference in unstable environments.

Together, these findings demonstrate that the model provides a robust advantage over a WSLS-based ToM heuristic, except at the highest learning rate. In this regime, rapid belief updating likely leads to overly strong beliefs that may not adapt well to highly volatile opponent behavior, resulting in inaccurate predictions. Importantly, the volatility levels used in these simulations reflect theoretical upper bounds and may not reflect typical dynamics in real-world social interactions. Full statistics for all comparisons, including means, standard errors, t-values, p-values, and Hedges' g, are reported in Supplementary Table 1.

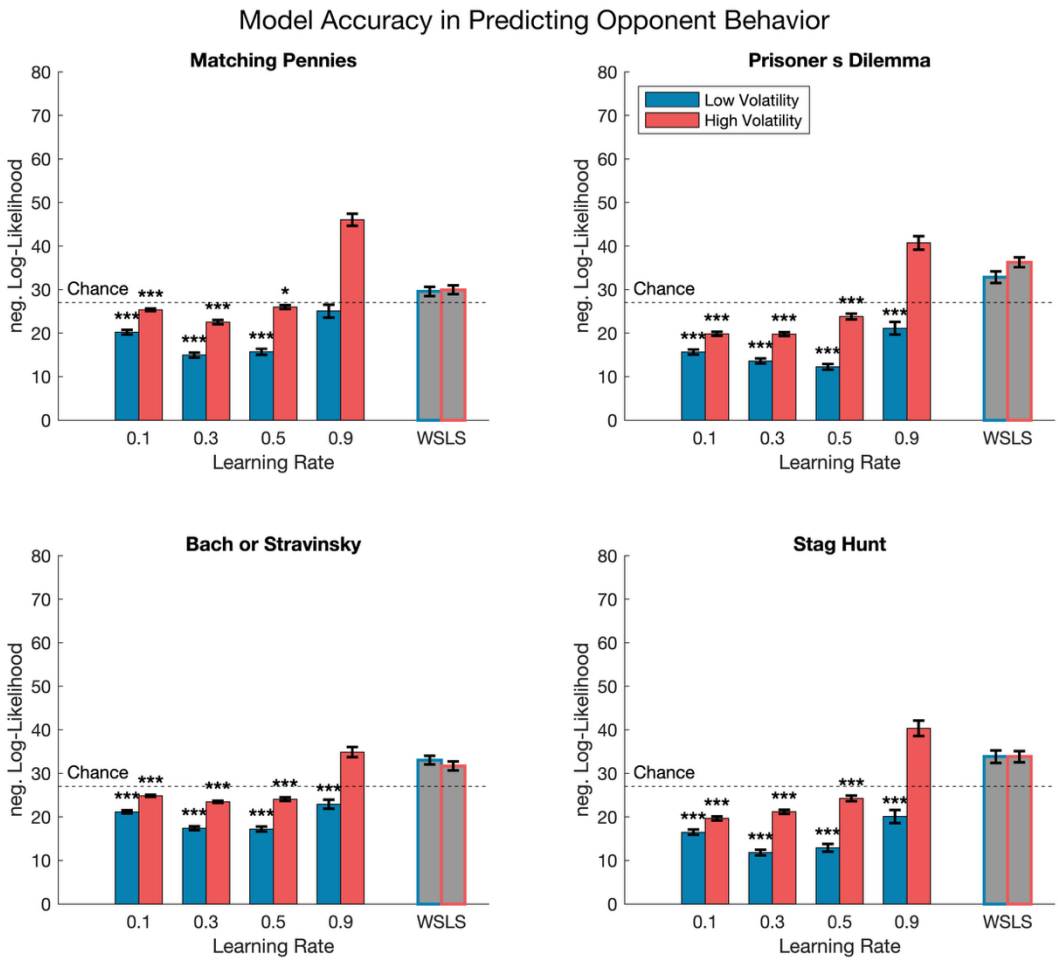

Figure 3 – Model Accuracy in Predicting Opponent Behavior; This figure illustrates the predictive accuracy of our ToM-based model across four classic dyadic games: Matching Pennies, Prisoner's Dilemma, Bach or Stravinsky, and Stag Hunt. Each game is simulated over 40 trials for 100 independent dyads. The model's learning rate parameter is systematically varied across four levels (0.1, 0.3, 0.5, 0.9) to assess how adaptive updating affects prediction under different conditions. We compare the model's ability to predict the opponent's next choice under low and high volatility conditions, defined by 10% and 40% probability of opponent strategy change per trial, respectively. Performance is benchmarked against both a classical Win-Stay-Lose-Shift (WSLS) model and chance level. The WSLS model deterministically predicts the opponent choice based on past action and reward. This predicted action is then converted into a probabilistic prediction using a softmax function with temperature = 1, ensuring direct comparability to the main model. Statistical significance is assessed using paired t-tests against chance performance. Asterisks in the figure indicate conditions where the model significantly outperforms chance ($p < 0.05$). This analysis highlights the robustness of the model across strategic contexts and volatility levels. In all panels, error bars reflect the SEM. ***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$ (one-tailed t-test, FDR corrected).

In Figure 4, we illustrate the belief dynamics over four L0 strategies, IMITATE, REPEAT, WSLS, and OPPOSE, in the Matching Pennies game under two volatility conditions: low (top subplot) and high (bottom panel). With a fixed learning rate of 0.5, the plots show how the posterior probability (y-axis) for each strategy evolves over time (x-axis), while the background color denotes the ground truth L0 strategy. This visualization reveals the underlying inference mechanism: as evidence accumulates, the posterior about the true strategy increases, governed by updates to the mean of a Dirichlet distribution (see Methods). These strategy-specific beliefs are then integrated to compute the expected opponent behavior at level 0 ($p(AG_0)$).

Notably, different latent strategies can produce indistinguishable observable behavior, making pure strategy classification inherently ambiguous in some contexts. For instance, from the perspective of a matcher in MP, WSLS and OPPOSE result always in the same observed actions. This overlap is evident in Figure 4, where the red and orange belief traces overlap consistently. Importantly, the model addresses this ambiguity by integrating across beliefs when computing the final opponent prediction, rather than relying on single-strategy identification. This underscores a fundamental limitation: environments with overlapping behavioral signatures impose structural constraints on how well any model can resolve latent strategies. However, this is not a shortcoming of the model itself. In fact, as shown below in Figure 5, while constrained below a perfect classification the model's performance approaches fair agreement beyond chance level.

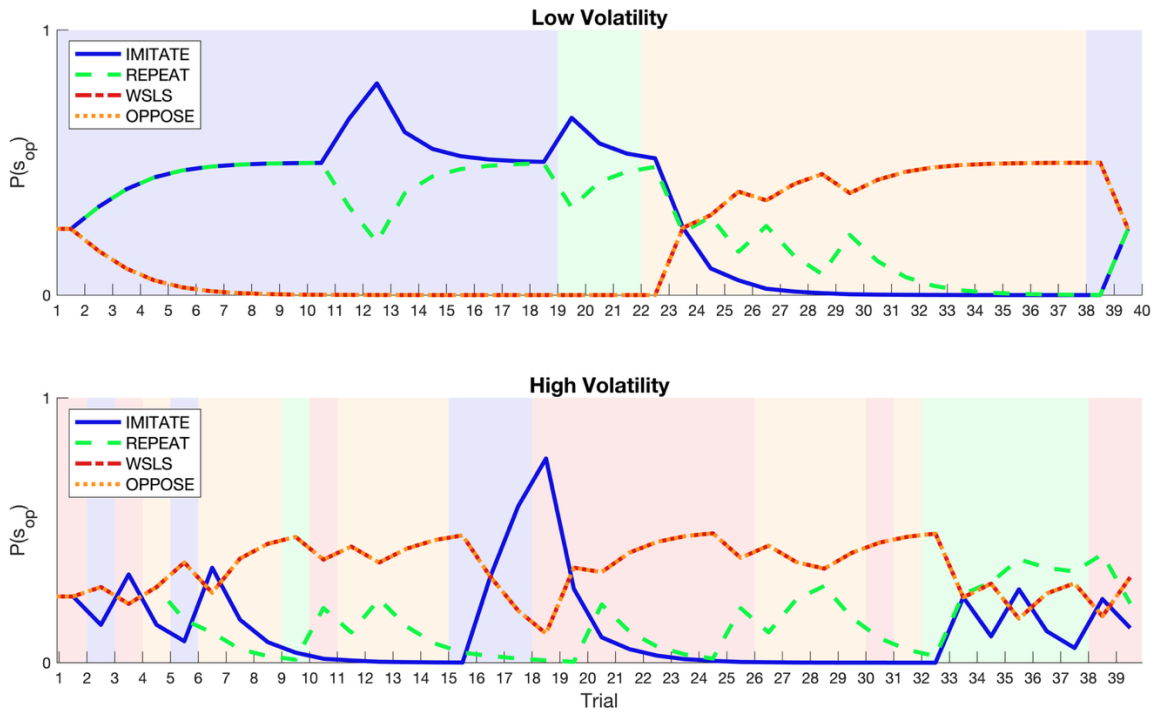

Figure 4 — Tracking strategy beliefs. Evolution of the model's beliefs over four level-zero (L0) strategies—IMITATE, REPEAT, WSLS, and OPPOSE—across 40 trials in the Matching Pennies game. The top panel shows behavior under low volatility (10% probability of switching strategy per trial), while the bottom subplot shows high volatility (40% switching probability). Colored background segments indicate the true underlying opponent strategy on each trial. Strategy belief traces reflect posterior probabilities over time, demonstrating how the model accumulates evidence and updates its beliefs accordingly.

To systematically evaluate the model's ability to infer the correct latent strategy, we computed a trial-by-trial confusion matrix by comparing the model's predicted strategy (maximum a posteriori) to the true L0 strategy. As a summary statistic, we used Cohen's Kappa (κ) (Cohen, 1960), which adjusts for chance agreement (κ = 0 indicates chance performance) and provides a conservative estimate of classification quality (cit.; see Methods and Figure 5). Figure 5 displays average κ values across simulated dyads on the four games, stratified by learning rate and volatility. To determine whether performance exceeded chance, we conducted one-sample t-tests comparing subject-level κ values against zero (chance level), correcting p-values for multiple comparisons using the FDR procedure. Confusion matrices were computed only for dyads in which more than one strategy was expressed over time; sequences containing a single, unchanging true strategy (e.g., 3.5% of MP game sequences) were excluded from the analysis.

Under low volatility (10% switching), the model reliably tracked the underlying L0 strategy across all games and learning rates. In all four games, κ values were significantly above zero at moderate to high learning rates (e.g., learning = 0.5: mean κ = 0.46, $t(94) = 17.89$, $p < 0.001$, Hedges' $g = 1.82$; learning = 0.9: mean κ = 0.44, $t(96) = 17.35$, $p < 0.001$, $g = 1.75$). Under high volatility (40% switching), the model's ability to infer the correct strategy was reduced. As evident for by the less significant p at learning rate 0.1 for Prisoner's Dilemma (mean κ = 0.03, $t(99) = 3.14$, $p = 0.002$, $g = 0.31$) and Stug Hunt (mean κ = 0.03, $t(99) = 3.00$, $p = 0.003$, $g = 0.30$).

Importantly, all positive Kappa values were significantly above chance, and values exceeding 0.2 are typically interpreted as reflecting at least fair agreement beyond chance in classification tasks (Landis & Koch, 1977). Under low volatility, the model's performance consistently approached or exceeded this threshold, indicating meaningful tracking of latent strategies. In contrast, under high volatility, although Kappa values remained above chance, they generally fell within the slight agreement range. This reduction is consistent with the model's architecture, which includes a one-trial delay in belief updating. Given that strategies can change on up to 40% of trials in the high-volatility condition, the model has limited opportunity to accumulate stable evidence before a new switch occurs.
These findings highlight a central trade-off in probabilistic inference: the model excels at tracking latent strategies under stable conditions but becomes less effective in volatile contexts due to delayed evidence accumulation.

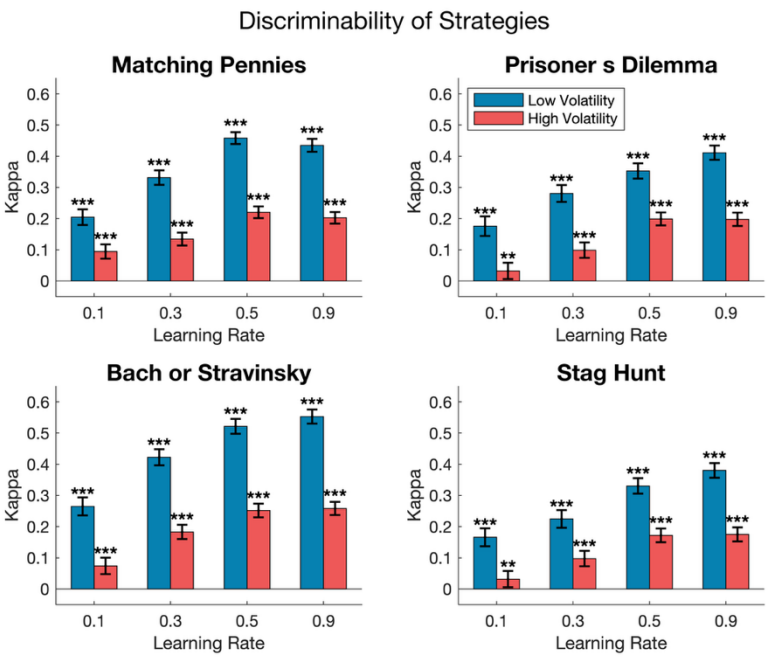

Figure 5 – Discriminability of Strategies: Average Cohen's Kappa values are shown across four games (MP, PD, BoS, SH), stratified by learning rate (x-axis) and volatility condition (bar color: blue = low, orange = high). Kappa quantifies agreement between the model's inferred L0 strategy and the ground truth, corrected for chance (κ = 0). Error bars represent ±SEM across participants. Asterisks denote significance against chance level: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$ (one-tailed t-test, FDR corrected).

Distinguishing opponent ToM levels and disentangling agents level

To assess whether the model can infer the recursive depth of an opponent's reasoning, we extended the simulation framework described above. For each game, we generated 40 action sequences and simulated 100 agent–opponent pairs across four learning rates (0.1, 0.3, 0.5, 0.9). As in previous analyses, volatility was manipulated by allowing the opponent's L0 strategy to switch with a 10%

(low) or 40% (high) probability on each trial. Critically, the opponent's ToM level (L0 or L1) was also allowed to switch independently at the same rate, enabling us to test whether the model could track the opponent's level in addition to their surface-level strategy. Moreover, the agent's actions were always computed from level L2. However, on every trial, BELIEFS estimates action probabilities for both opponent's level L0 and L1 (given by $p(OP_0)$ and $p(OP_1)$) as well as agent's level L0, L1 and L2 (given by $p(AG_0)$, $p(AG_1)$ and $p(AG_2)$).

To quantify inference quality, we computed the average negative log likelihood (NLL) of the opponent's choices given the model's prediction $p(OP_0)$ on trials where the opponent's true level was L0 (Figure 6, top left), and compared it to the NLL of $p(OP_0)$ on trials where the opponent was actually L1 (Figure 6, top right). Similarly, we computed the average NLL of $p(OP_1)$ for trials where the opponent's true level was L1 (Figure 6, bottom left), and compared it to the NLL of $p(OP_1)$ on trials where the opponent was actually L0 (Figure 6, bottom right).

Across all games and conditions, the model reliably assigned lower NLL to actions belonging the true ToM level than to the incorrect one. For example, in the Matching Pennies (MP) game, true ToM0 agents were associated with significantly lower NLL than ToM1 (mean NLL difference = 0.61, one-tailed $t(7) = 3.24$, $p = 0.007$, Hedges' $g = 1.10$) and vice versa for ToM1 opponents (but see learning rate of 0.1 at low volatility for MP). Figure for results in PD, BoS, and SH are in Supplementary Material). The model's inference was consistently significantly below chance, indicating robust sensitivity to the opponent's reasoning depth, especially under lower volatility.

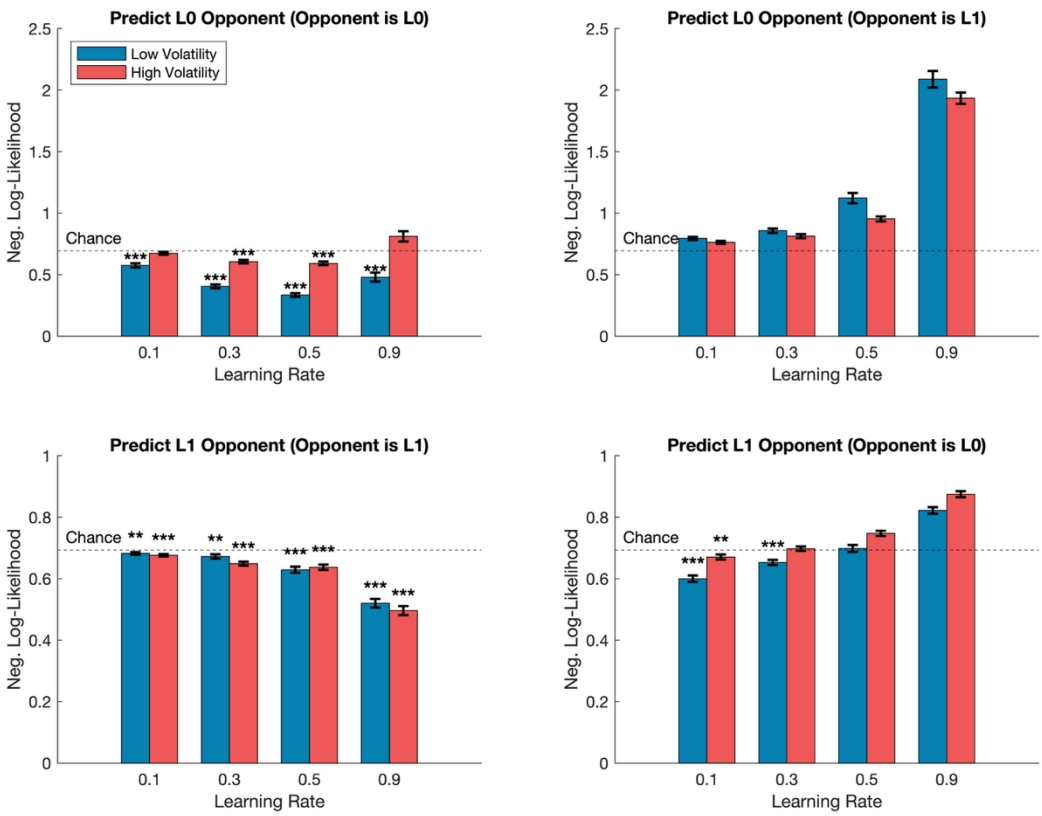

Figure 6. To assess BELIEFS' ability to infer opponent's level of reasoning sophistication, we compared trial-by-trial negative log-likelihoods (NLL) of opponent's expected action, given the actual taken, between the true ToM level (left column) and the incorrect level (right column). In the Matching Pennies (MP) game, true ToM0 agents were associated with significantly lower NLLs than ToM1, indicating better model alignment with their actual reasoning level (mean NLL difference = 0.61, one-tailed $t(7) = 3.24$, $p = 0.007$, Hedges' $g = 1.10$). The inverse pattern was observed for ToM1

opponents. See Supplementary Figure for analogous results in the PD, BoS, and SH games. Error bars indicate ±SEM across participants. Asterisks denote significance against chance level: **p < 0.001, *p < 0.01, p < 0.05 (one-tailed t-test, FDR corrected).

Next, we evaluated the model's ability to disentangle the agent's own ToM level, a key feature for fitting behavioral data. Specifically, we measured the dissimilarity between model-generated action probabilities for agents at ToM levels 0 ($p(AG_0)$), 1 ($p(AG_1)$), and 2 ($p(AG_2)$), using Fisher-transformed correlations across trials. Greater dissimilarity reflects more distinctive behavioral profiles across levels, supporting reliable inference of the agent's reasoning depth.

As shown in Figure 7, dissimilarity scores were significantly greater than zero for most pairwise comparisons ($p < 0.001$, one-tailed t-tests against 0, FDR-corrected), particularly in competitive games. However, in cooperative tasks (BoS and SH), agents at level 0 and level 2 often produced highly correlated behaviors, resulting in lower, but still significant, dissimilarity. This suggests that cooperative contexts may reduce the distinguishability of ToM levels, as different agents may converge on similar strategies despite differing reasoning processes.

Together, these results show that the model can accurately infer both the opponent's and the agent's level of recursive reasoning. This dual capacity to track latent opponent ToM levels and differentiate between internally generated reasoning levels is a key strength of the model, particularly under stable, competitive conditions.

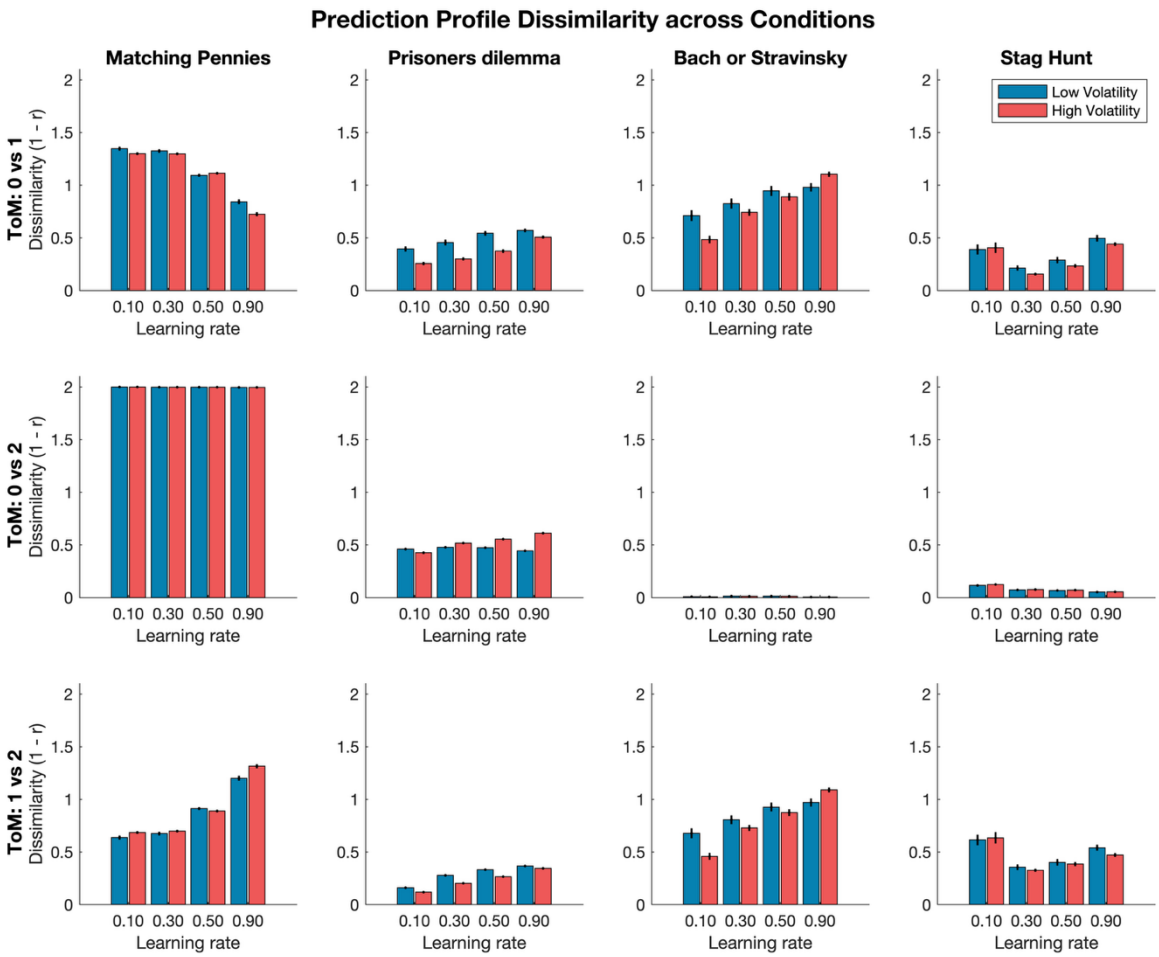

Figure 7. To assess the BELIEFs' ability to distinguish between different Theory of Mind (ToM) levels, we computed the dissimilarity between trial-wise action probability distributions for agents at ToM levels 0, 1, and 2 (respectively, $p(AG_0)$, p(AG1), and $p(AG_2)$). This dissimilarity reflects the theoretical separability of reasoning levels based on observed behavior across four matrix games. In the Matching Pennies (MP) game, dissimilarity scores were consistently high—indicating strong behavioral distinction between levels—with ToM0 and ToM2 agents exhibiting opposite (anticorrelated) actions, as reflected in dissimilarity values approaching 2. In contrast, for cooperative games such as the Prisoner's Dilemma (PD), Bach or Stravinsky (BoS), and Stag Hunt (SH), dissimilarities—particularly between ToM0 and ToM2—were notably lower, suggesting convergence on similar strategies despite differences in underlying reasoning. Notably, in BoS, ToM0 and ToM2 agents produced nearly indistinguishable behavior. Nonetheless, all pairwise comparisons were significantly greater than zero ($p < 0.001$, FDR corrected), indicating possible differentiation across ToM levels. Error bars represent ±SEM across participants.

Discussion

Theory of Mind (ToM) is not a unitary process (Rusch, Steixner-Kumar, Doshi, Spezio, & Gläscher, 2020), but rather a family of socio-cognitive mechanisms that support the attribution of first-order beliefs and values to others, integrating these representations into one's own decision-making. In this work, we present a model that advances current approaches by relaxing a key constraint: the fixed Level-0 (L0) strategy assumption, which is often instantiated using the Win-Stay-Lose-Shift (WSLS) heuristic (Devaine et al., 2014; Yoshida et al., 2008). Instead, our framework allows for probabilistic inference over a broader set of opponent strategies, enhancing both the accuracy of ToM-level estimation and the predictive validity of the model. By moving beyond rigid heuristics such as WSLS and adopting a distributional approach to strategy and ToM-level inference, the model better captures how humans might reason about others. Prior models frequently rely on rational agent assumptions that presume optimality or fixed rules, assumptions often violated in real-world behavior (Arthur, 2021; Erev & Roth, 2014). In contrast, our framework embraces behavioral variability and supports bounded rationality by enabling agents to infer, weight, and adapt to a range of possible opponent strategies and reasoning depths.

Another key limitation in existing ToM models is their restricted generalizability across different task contexts and their fixed assumption regarding the level of mentalization. Our model addresses this by incorporating ongoing updates to beliefs and strategies, allowing it to flexibly adjust to changing opponent behavior. This feature broadens the model's applicability beyond constrained laboratory settings to more ecologically valid and dynamic environments. Nevertheless, further behavioral validation is necessary to assess how well these mechanisms generalize beyond simulations. A central strength of the proposed architecture is its scalability. It can, in principle, be applied to any matrix game structure, offering a marked improvement over models that fail in more complex, multi-agent, or time-sensitive scenarios. That said, the model's performance deteriorates under conditions of high volatility; where environmental changes outpace belief updates. This is an expected limitation, as extreme volatility represents a theoretical edge case where belief tracking naturally becomes unstable due to persistent misalignment between observed behavior and belief updating.
While our model offers a promising framework for scalable and cognitively plausible ToM reasoning, it has thus far been evaluated only on simulated data. Future research should test its robustness against empirical behavioral datasets, incorporate prediction error signals into the belief-updating dynamics, and explore alternative formulations of the concentration parameters (α). It will also be important to examine how social reasoning interacts with affective or moral components of cognition; currently outside the scope of our model.

Notably, our framework is structured to support inference over both the agent's and the opponent's ToM levels, enabling the tracking of recursive beliefs, not only about others' strategies but also about their reasoning depth. While not fully implemented in the current analyses, the model architecture

allows for maintaining and updating probabilistic beliefs over different ToM levels based on the relative predictive accuracy at each level. This opens a promising direction for future work, particularly in addressing long-standing debates about whether human mentalizing operates via discrete levels or more continuous gradients of belief. Moreover, such probabilistic outputs could, in principle, be aligned with neural data (e.g., functional imaging) to study how social representations evolve over time during decision-making. In future work, we aim to more directly evaluate the model's inference over ToM levels by computing trial-by-trial posterior probabilities of each level given the opponent's action.

Beyond theoretical implications, flexible ToM tracking has substantial applications. In human-AI interaction, systems that can infer human goals and social reasoning depth can respond more adaptively in collaborative and mixed-initiative contexts, such as assistive robotics or digital companions (Langley et al., 2022). Similarly, in education, such models could underpin adaptive learning systems that infer student misunderstandings not only from observable errors but also from inferred reasoning strategies and beliefs (Baker, Saxe, & Tenenbaum, 2009). In both domains, dynamic and probabilistic ToM modeling bridges computational inference with the variability and nuance of human cognition.

In summary, we introduce a novel, flexible model of Theory of Mind that accumulates evidence over multiple strategies and integrates hierarchical beliefs to compute best-response action probabilities. By relaxing rigid prior assumptions and dynamically tracking social reasoning depth, the model provides a scalable and cognitively informed foundation for future research in decision science, neuroscience, and human-AI interaction.

.

Material and Methods

Software implementation and availability

All codes and simulation data are available for download on GIN.
All data simulation and analyses were carried out in Matlab 2024a (The Mathworks)
Statistics for p-value correction was done with fdr_bh toolbox for matlab («fdr_bh», 2025) following the (Benjamini & Hochberg, 1995) method .

L0 Strategies definition and set choice

Level-0 (L0) strategies in the model are defined as simple heuristics that generate deterministic predictions based on recent contextual information; namely, the most recent actions and outcomes observed in the game. These strategies operate without any Theory of Mind (ToM) reasoning and treat the opponent's behavior as an environmental regularity rather than the product of intentional decision-making. In the present study, we consider a fixed set of four candidate L0 strategies commonly discussed in the literature on repeated matrix games (e.g., Matching Pennies, Prisoner's Dilemma; (Du, Wu, Altrock, & Wang, 2014; Imhof, Fudenberg, & Nowak, 2007; Ioannou & Romero, 2014)). However, it is important to note that the model is not limited to this set—additional strategies can be included or excluded depending on the context of application. The four strategies used are defined as follows:

IMITATE. This strategy predicts that the opponent will repeat the agent's last choice ($c_{OP}(t+1) = c_{AG}(t)$). In the literature on the Prisoner's Dilemma game, this strategy is known as Tit-for-Tat.

REPEAT. This strategy predicts that the opponent is simply repeating his last choice ($c_{OP}(t+1) = c_{OP}(t)$). This strategy partially overlaps with the winning part of WSLS.

OPPOSE. *WSLS (Win-Stay, Lose-Shift).* This strategy predicts that the opponent will repeat his current choice following a win in the current trial ($c_{OP}(t+1) = c_{OP}(t) \mid win$), whereas he will shift to the other choice option following a loss in the current trial ($c_{OP}(t+1) \neq c_{OP}(t) \mid loss$).

This strategy is the opposite of IMITATE and predicts that the opponent will oppose the agent and deliberately not choose the agent's last choice ($c_{OP}(t+1) \neq c_{AG}(t)$)

Calculating action probabilities at Level 1

The two players in this model are referred to as the agent (AG) and her opponent (OP). Modeling is always performed from the perspective of the agent with a particular ToM level *k* ($k \geq 1$) who is building a mental model of her opponent at level *k-1*. This section specifies the calculations at Level 1.

Each player selection actions at trial, which are denoted c_{AG} for the agent and c_{OP} for the opponent. These actions are evaluated against a finite set of strategies $S = \{s_1, \ldots, s_n\}$ (e.g. $S = \{IMITATE, REPEAT, WSLS, OPPOSE\}$; see above). The (binary) *strategy-specific belief* over choices distribution $b_0(c_{OP}|s)$ of L0 strategies (defined in more detail below) specifies the probability of the current opponent choice $p(c_{OP})$ under a specific strategy s.

The model calculates action probabilities for the agent using a standard expected value approach ($EV = x \times p(x)$) (Suri, Gross, & McClelland, 2020), where x is the outcome and $p(x)$ is the probability of the occurrence of x. However, in social interactive matrix games, the outcome (utility) depends on the choices of both players $U(c_{AG}, c_{OP})$. Given that the agent chooses c_{AG}, she will only get the outcome $U(c_{AG}, c_{OP})$ with the probability $p(c_{OP})$, namely that the opponent chooses his current choice c_{OP}. At Level 1, when the agent is observing the L0 opponent and detecting, which strategy he is executing, his choice probability $p_0(c_{OP})$ is defined by the integrated belief over all strategies $B_0(c_{OP})$ (see below for details on the belief integration).

Thus, the L1 agent computes the values of her two choices using:

$V_1(c_{AG}) = B_0(c_{OP}) \times U(c_{AG}, c_{OP})$,

where $V_1(c_{AG})$ is an *n*-element vector of expected values for the agent's choice options (calculated for a L1 agent), $B_0(c_{OP})$ is an *n*-element vector of integrated beliefs about the L0 opponent's choice options (marginalizing all specific strategy *s*, see below for details) and $U(c_{AG}, c_{OP})$ are the agent's outcomes (utilities) given the joint choice (c_{AG}, c_{OP}). Note that $U(c_{AG}, c_{OP})$ can be different for the agent and the opponent, which is why we use $U(c_{AG}, c_{OP})$, when we mean the agent's payoffs, and $U(c_{OP}, c_{AG})$ when referring to the opponent's payoffs.

By calculating action values for every L1 agent's choice option the model is marginalizing the L0 opponent's choice and deriving the value for the agent's best response. Indeed, if we spell out the above mentioned equation for each of two options ($c_{AG} = \{1,2\}$) of the L1 agent, we get the following equations:

$$V_1(c_{AG} = 1) = B_0(c_{OP} = 1) \times U(c_{AG} = 1, c_{OP} = 1) + B_0(c_{OP} = 2) \times U(c_{AG} = 1, c_{OP} = 2)$$

$$V_1(c_{AG} = 2) = B_0(c_{OP} = 1) \times U(c_{AG} = 2, c_{OP} = 1) + B_0(c_{OP} = 2) \times U(c_{AG} = 2, c_{OP} = 2)$$

The joint outcomes $U(c_{AG}, c_{OP})$ (and $U(c_{OP}, c_{AG})$) are usually known to both players, so the only unknown in the computation of expected value is $B_0(c_{OP})$, the integrated beliefs about the opponent's next action. The main goal of the model (and indeed of most other recursive ToM models (de Weerd et al., 2013; Devaine et al., 2014; Yoshida et al., 2008) is to estimate this belief probability of the opponent's actions. The current model departs from its predecessors by integrating the action probability under different L0 opponent strategies of the opponent with the beliefs about these strategies (see below).

Finally, the expected values $V_1(c_{AG})$ are then filtered through a softmax choice function to derive action probabilities $p_1(c_{AG})$ for the agent:

$$p_1(c_{AG}) = e^{\tau V_1(c_{AG})} / \sum_{c_{AG}} e^{\tau V_1(c_{AG})}$$

Where τ is a free temperature parameter modeling the stochasticity of the choice (i.e. how much the action probability adheres to the distance between action values. We will abbreviate this equation as $p_1(c_{AG}) = softmax(\tau V_1(c_{AG}))$. In this study, temperature was fixed to 1.

L0 belief integration

The model for an L1 agent cycles over a finite set of strategies $s \in S$ and evaluates whether the current choice of the L0 opponent matches each strategy and accumulates for it.

Currently, all strategies evaluate past choices and outcomes to derive a discrete choice prediction for the current trial, also known as a (binary) strategy-specific belief over choices. These take the form $b_0(c_{OP}|s) = [1\ 0]$ if $c_{OP} = 1$ or $b_0(c_{OP}|s) = [0\ 1]$ if $c_{OP} = 2$. For instance, the *IMITATE* strategy predicts that $c_{OP}(t) = c_{AG}(t-1)$, so if $c_{AG}(t-1) = 1$, then $b_0(c_{OP}(t)|s = IMITATE) = [1\ 0]$. However, the prediction of the same strategy for different trials may not be identical: the opponent's choice $c_{OP}(t+1)$ on the next trial may be different (though still consistent with the IMITATE strategy), because the agent also chose a different action on trial t ($c_{AG}(t) \neq c_{AG}(t-1)$). In this case, the $b_0(c_{OP}(t+1)|s = IMITATE)$ belief will shift from, e.g. [1 0] to [0 1] on the next trial.

To accumulate evidence for each strategy across trials, the model maintains and updates a (Dirichlet) belief distribution over strategies:

$$p(s) \sim Dir(\alpha_s)$$

Where α_s is the strategy-specific Dirichlet parameter that counts the matches with the strategy. For example, in the simplest form of Dirichlet updating, the α_s are increased by 1, whenever the respective strategy is detected, e.g. if $b_0(c_{OP}(t)|s = IMITATE) = [1\ 0]$ and $c_{OP}(t) = 1$, then $\alpha_{IMITATE} = \alpha_{IMITATE} + 1$ (but see below "*update rule for s*").

The binary strategy-specific beliefs of choices $b_0(c_{OP}|s)$ and the hierarchical belief distributions over strategies $p(s)$ are then integrated into the single overall belief distribution $B_0(c_{op})$ by

$$B_0(c_{OP}) = \sum_{s} p(s)\, b_0(c_{op}|s)$$

This integrated belief over the opponent's choices is then used to calculate the agent's values $V_0(c_{AG}) = B_0(c_{OP}) \times U(c_{AG}, c_{OP})$ (see above).

Note that the player that the L0 beliefs are referring to can change depending on the current ToM level of the agent: all odd-numbered ToM levels agent are built on the L0 strategies of the opponent (and the best responses to them), whereas all even-numbered ToM levels are built on the L0 strategies of the agent (as observed by the L1 opponent. Thus, the model must maintain and update two different Dirichlet distributions over L0 strategies for each L0 player $p_{OP}(s) \sim Dir_{OP}(\alpha_s)$ and $p_{AG}(s) \sim Dir_{AG}(\alpha_s)$.

Belief Updating for L0 strategies

In the proposed model, the agent maintains a belief distribution over a finite set of Level-0 (L0) strategies $S = \{s_1, , s_n\}$ which correspond to possible rules the opponent might be using (e.g., Win-Stay-Lose-Shift, Imitate, Oppose). To represent uncertainty over strategies, the model employs a Dirichlet distribution, parameterized by a concentration vector $\alpha = \alpha_s$, with one parameter α_s, for each strategy s in S. The agent's belief over strategies is given by the mean of this Dirichlet distribution:

$$P \sim Dir(\alpha_s)$$
$$E[P(s)\ \ ] = \frac{\alpha_s}{\sum_s \alpha_s{}'}$$

This formulation captures both the probability and the confidence (or precision) in each strategy: as α_s increases, the belief in strategy s becomes stronger and more certain. The Dirichlet framework ensures that the resulting belief distribution is valid—non-negative and summing to one. For simplicity, the model uses the mean of the Dirichlet as a point estimate of the agent's belief over strategies

However, traditional Dirichlet updates follow a simple "learning by counting" scheme in which α_s is incremented upon each match (e.g., $\alpha_s \leftarrow \alpha_s + 1$ implicitly assuming that the opponent's strategy is fixed throughout the interaction. Under this scheme, beliefs become increasingly peaked over time, making it difficult to adapt when the opponent changes strategies. As evidence accumulates, the influence of new observations diminishes, and the model adapts too slowly, particularly in dynamic settings where strategies may shift.
To address this limitation, we introduce a flexible update rule that incorporates adaptive forgetting. Rather than simply incrementing counts, the concentration parameters are updated using a decayed evidence accumulation mechanism:

$$\alpha_s \leftarrow (1 - lr)\ \times \alpha_s + lr \times b(c_{OP}|s)$$

Here, $lr\ \in [0,1]$ is a fixed learning rate that regulates how strongly the new evidence influences the update, and $b(c_{OP}|s) = 1$ is a binary indicator: it equals 1 if the opponent's choice c_{OP} matches the prediction of strategy s, and 0 otherwise (please note that $b(c_{OP}|s)$ represents the detection of strategies (matches) while the previously reported $b_0(c_{OP}|s)$ indicates the deterministic prediction binary outcome given strategy s). This update discounts all strategies by a of $(1 - lr)$, allowing their weights to decay over time unless reinforced by matching observations. The mechanism enables

both belief strengthening and weakening, thus supporting rapid adaptation to shifts in the opponent's behavior.

Belief integration and valuation at ToM level 2

At ToM level 2 the agent simulates a mental model of the opponent as a ToM level 1 player, which includes a simulation of the agent as a Level-0 player extracting strategies. Note that all of these simulations at different levels are part of the integrated model of a ToM-2 agent. As in the simulations at Level-2 the recursive ToM hierarchy is always constructed from bottom-up starting with the Level-1 player who is observing the Level-0 player and then iteratively best-responding to the lower level predictions.
The (ToM-2)-simulated level 1 agent is integrating their observations of the opponent's choices and strategies into an integrated belief

$$B_0(c_{AG}) = p(s) \odot b_0(c_{AG}|s)$$

from which expected values and simulated action probabilities for the ToM-0 agent are computed:

$$V_1(c_{OP}) = B_0(c_{AG}) \times U(c_{AG}, c_{OP})$$

$$p_1(c_{OP}) = softmax(\tau V_1(c_{OP}))$$

Note that this is the same belief integration as in the first case of a ToM-1 agent, but now the ToM-2 agent extends this simulation by computing expected values and action probabilities for the simulated ToM-2 opponent, who is best-responding by marginalizing the simulated Level-1 agent :

$$V_2(c_{AG}) = p_1(c_{OP}) \times U(c_{AG}, c_{OP})$$

$$p_2(c_{AG}) = softmax(\tau V_2(c_{AG}))$$

Extending the hierarchy beyond Level 2

To summarize the construction of the Theory of Mind (ToM) hierarchy, at Level 1 (L1), the opponent is modeled as an L0 player whose choices and strategies are observed and integrated into a belief distribution over their possible actions. Based on this belief distribution, expected values for the agent's actions are computed by marginalizing over the opponent's predicted choices. These expected values are then transformed into action probabilities using a softmax function. At the next higher ToM level (L2), the process is repeated: the agent uses the previously derived action probabilities to marginalize over the opponent, now considered an L1 player, to compute new expected values and corresponding action probabilities for the L2 level. This recursive mechanism continues iteratively for higher ToM levels. Consequently, a general formulation for any ToM level k > 0 can be expressed through the recursive application of these steps, where k denotes the ToM level of the agent constructing the model:

$$V_k(c_{AG}) = p_{k-1}(c_{OP}) \times U(c_{AG}, c_{OP})$$

$$p_k(c_{AG}) = softmax(\tau V_k(c_{AG}))$$

Prediction opponent choice analysis

To evaluate model performance in predicting opponents' choices across different learning rates and volatility conditions, we computed the negative log-likelihood (NLL) of model-based predictions. For

each combination of learning parameter and volatility (low vs. high), we used simulated model beliefs to estimate the probability of the opponent's actual choice on each trial. The model's predicted probability for the observed opponent's action was logged (log-likelihood), and summed across trials (excluding the first) to produce a subject-wise total log-likelihood score.

We compared this model-based log-likelihood against both chance level performance and a Win-Stay Lose-Shift (WSLS) heuristic. The WSLS model deterministically predicts the opponent will repeat their previous choice if they won (i.e., obtained an equal or better outcome than the agent) and switch otherwise. This predicted action was then converted into a probabilistic prediction using a softmax function with temperature = 1, ensuring direct comparability to the main model. The softmax temperature of 1 corresponds to an effective transformation factor of 0.269 applied to the WSLS predictions.

For both the main model and the WSLS benchmark, individual trial-wise log-likelihoods were computed, and group-level mean log-likelihoods were derived. We conducted one-sided t-tests comparing each model's log-likelihoods to chance level, and paired t-tests comparing the main model against WSLS. In all cases, we also report Hedges' g as an effect size metric, correcting for sample size.

This approach allowed us to systematically assess how model performance varies as a function of internal learning parameters and environmental volatility, and whether it reliably outperforms a baseline heuristic.

Strategies confusion matrix analysis

To evaluate how well the model discriminates among latent strategies, we computed Cohen's kappa (Cohen, 1960) between the model-predicted and true strategies across trials. This measure corrects for chance agreement and provides a normalized metric of classification performance, with $\kappa = 1$ indicating perfect prediction and $\kappa = 0$ corresponding to chance-level agreement.

For each trial, the model outputs a posterior probability distribution over strategies, denoted as $P(s)$, where s indexes the strategy space. To generate discrete predictions, we selected the strategy with the maximum posterior. In cases of ties (i.e., multiple strategies with equal maximum probability), we resolved them by randomly sampling one of the tied strategies. To account for the stochasticity introduced by random tie-breaking, we repeated this procedure 100 times and computed a confusion matrix and kappa coefficient for each sample. The final model-based kappa was obtained by averaging across bootstraps.

For each subject, we constructed a normalized confusion matrix by comparing predicted and true strategies across trials. Rows were normalized to sum to one, representing conditional probabilities of predicting each strategy given the true label. We then averaged these confusion matrices across subjects and bootstrap samples to compute the overall confusion matrix. Kappa was computed from this averaged confusion matrix using standard formulae for multiclass classification.

Opponent level prediction

To assess the model's ability to infer the opponent's level of strategic sophistication across varying learning rates and environmental volatility, we computed the negative log-likelihood (NLL) of the model-based predictions under four diagnostic conditions. For each participant and parameter configuration, the probability assigned by the model to the opponent's actual choice on each trial was estimated using either a level-0 (B0) or level-1 (B1) opponent model. These predicted probabilities were log-transformed, and the resulting log-likelihoods were averaged across trials (excluding the initial trial) and then across subjects.

To evaluate both correct and incorrect inferences, log-likelihoods were computed under four distinct scenarios. First, prediction accuracy was assessed when the opponent truly employed a level-0

strategy and the model generated a level-0 prediction (B0-true). Second, predictions were evaluated when the opponent used a level-0 strategy but the model assumed a level-1 strategy (B0-false). Third, prediction accuracy was measured when the opponent genuinely used a level-1 strategy and the model applied a level-1 prediction (B1-true). Fourth, predictions were examined when the opponent was level-1 but the model assumed a level-0 strategy (B1-false).
For each of these four metrics, the average log-likelihood across subjects was compared against chance performance, defined as log(0.5), using one-sided t-tests to determine whether the model's predictions significantly exceeded chance levels. Additionally, the standard error of the mean and Hedges' g, as a bias-corrected effect size, were reported.
This evaluation was repeated for all combinations of learning rate and volatility (hazard rate), enabling quantification of the model's reliability in distinguishing the opponent's true reasoning level under different environmental dynamics.

Disentangling agent levels

We quantified the dissimilarity between model predictions across ToM levels by computing the pairwise dissimilarity of the action-wise log-likelihoods predicted by the model under different ToM levels (0, 1, and 2). For each simulated agent, we constructed a 3×3 Fisher-z-transformed correlation matrix between the predicted action log-likelihoods for each pair of ToM levels. These correlation matrices were averaged across trials and then converted into dissimilarity values as:

$$Dissimilarity = 1 - tanh(z)$$

where z is the Fisher-z-transformed correlation between two ToM levels. We extracted the dissimilarity for each pair (0 vs 1, 0 vs 2, 1 vs 2) and analyzed how it varied across model parameters (learning rates) and task conditions (volatility).

Acknowledgment

This work was funded by the German Research Council DFG, Collaborative Research Center SFB 1528 "Cognition of Interaction".

CRediT authorship contribution statement

G.L.M; Conceptualization, Writing - original draft, Software, Analysis, Visualization. S.K.; Conceptualization. O.S.; Conceptualization, Writing - original draft. J G; Conceptualization, Funding acquisition, Writing - original draft, Software.

References

Arthur, W. B. (2021). Foundations of complexity economics. *Nature Reviews Physics*, *3*(2), 136–145. https://doi.org/10.1038/s42254-020-00273-3
Axelrod, R. (1980). More Effective Choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, *24*(3), 379–403. https://doi.org/10.1177/002200278002400301
Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005
Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A Cognitive Hierarchy Model of Games*. *The Quarterly Journal of Economics*, *119*(3), 861–898. https://doi.org/10.1162/0033553041502225
Camerer, C., Ho, T., & Chong, K. (2003). Models of Thinking, Learning, and Teaching in Games. *American Economic Review*, *93*(2), 192–195. https://doi.org/10.1257/000282803321947038
Cho, I., Kamkar, N., & Hosseini-Kamkar, N. (2022). Reasoning about mental states under uncertainty. *PLOS ONE*, *17*(11), e0277356. https://doi.org/10.1371/journal.pone.0277356
Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104
De Weerd, H., Diepgrond, D., & Verbrugge, R. (2018). Estimating the Use of Higher-Order Theory of Mind Using Computational Agents. *The B.E. Journal of Theoretical Economics*, *18*(2). https://doi.org/10.1515/bejte-2016-0184
de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, *199–200*, 67–92. https://doi.org/10.1016/j.artint.2013.05.004
Devaine, M., Hollard, G., & Daunizeau, J. (2014). Theory of Mind: Did Evolution Fool Us? *PLoS ONE*, *9*(2), e87619. https://doi.org/10.1371/journal.pone.0087619
Du, J., Wu, B., Altrock, P. M., & Wang, L. (2014). Aspiration dynamics of multi-player games in finite populations. *Journal of The Royal Society Interface*, *11*(94), 20140077. https://doi.org/10.1098/rsif.2014.0077
Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences*, *111*(supplement_3), 10818–10825. https://doi.org/10.1073/pnas.1402846111
Esterwood, C., & Robert, L. P. (2023). The theory of mind and human–robot trust repair. *Scientific Reports*, *13*(1), 9877. https://doi.org/10.1038/s41598-023-37032-0
Estes, D., & Bartsch, K. (2017). Theory of mind: A foundational component of human general intelligence. *Behavioral and Brain Sciences*, *40*, e201. https://doi.org/10.1017/S0140525X16001618
Fdr_bh. (2025, giugno 16). Recuperato 16 giugno 2025, da https://de.mathworks.com/matlabcentral/fileexchange/27418-fdr_bh
Imhof, L. A., Fudenberg, D., & Nowak, M. A. (2007). Tit-for-tat or Win-stay, Lose-shift? *Journal of theoretical biology*, *247*(3), 574–580. https://doi.org/10.1016/j.jtbi.2007.03.027
Ioannou, C. A., & Romero, J. (2014). Learning with repeated-game strategies. *Frontiers in Neuroscience*, *8*, 212. https://doi.org/10.3389/fnins.2014.00212
Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLOS ONE*, *3*(7), e2597. https://doi.org/10.1371/journal.pone.0002597
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
Langley, C., Cirstea, B. I., Cuzzolin, F., & Sahakian, B. J. (2022). Theory of Mind and Preference Learning at the Interface of Cognitive Science, Neuroscience, and AI: A Review. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.778852
Montes, N., Luck, M., Osman, N., Rodrigues, O., & Sierra, C. (2023). Combining theory of mind and abductive reasoning in agent-oriented programming. *Autonomous Agents and*

Multi-Agent Systems, *37*(2), 36. https://doi.org/10.1007/s10458-023-09613-w
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512
Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang, C., Eslami, S. M. A., & Botvinick, M. (2018, marzo 12). *Machine Theory of Mind*. arXiv. https://doi.org/10.48550/arXiv.1802.07740
Rusch, T., Steixner-Kumar, S., Doshi, P., Spezio, M., & Gläscher, J. (2020). Theory of mind and decision science: Towards a typology of tasks and computational models. *Neuropsychologia*, *146*, 107488. https://doi.org/10.1016/j.neuropsychologia.2020.107488
Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, *162*, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018
Stahl, D. O. (1993). Evolution of Smartn Players. *Games and Economic Behavior*, *5*(4), 604–617. https://doi.org/10.1006/game.1993.1033
Suri, G., Gross, J. J., & McClelland, J. L. (2020). Value-based decision making: An interactive activation perspective. *Psychological Review*, *127*(2), 153–185. https://doi.org/10.1037/rev0000164
Traulsen, A., Semmann, D., Sommerfeld, R. D., Krambeck, H.-J., & Milinski, M. (2010). Human strategy updating in evolutionary games. *Proceedings of the National Academy of Sciences*, *107*(7), 2962–2966. https://doi.org/10.1073/pnas.0912515107
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5
Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game Theory of Mind. *PLoS Computational Biology*, *4*(12), e1000254. https://doi.org/10.1371/journal.pcbi.1000254
Zhen, S., & Yu, R. (2021). Neural correlates of recursive thinking during interpersonal strategic interactions. *Human Brain Mapping*, *42*(7), 2128–2146. https://doi.org/10.1002/hbm.25355