A new multi-level Theory of Mind model for strategic decision-making scenarios

Giorgio Manenti (g.manenti@uke.de);
Sepideh Khoneiveh (s.khoneiveh@uke.de);
Jan Gläscher (glaescher@uke.de)
Department of Systems Neuroscience, University Medical Center Eppendorf Hamburg 20246 Hamburg, Germany

Abstract

Theory of Mind (ToM) enables individuals to infer others' intentions and beliefs. While existing models successfully simulate recursive ToM reasoning, they typically rely on a fixed strategy at the lowest level (e.g., Win-Stay-Lose-Shift). We introduce a novel ToM model that combines a dynamic belief over multiple strategies with recursive best responses at higher levels. Using simulated data from a Matching Pennies game, we show that the model accurately recovers underlying strategies, adapts to changes in the environment, and correctly infers the ToM level of the agent. This approach offers a more flexible and robust framework for modeling strategic social reasoning and opens new directions for understanding decision-making in interactive settings.

Keywords: Theory of Mind; dynamic belief updating; social decision making

Introduction

Theory of Mind (ToM) refers to the ability to infer others' latent cognitive states, such as intentions, goals, or beliefs. In recursive ToM, individuals model what others believe about them (e.g., "I think you think I'll do X"). This recursive reasoning is organized into levels, beginning with Level-0, where the agent acts without modeling the opponent, and assuming that others reason at one level below.

Recursive ToM models capture both competitive (Matching Pennies (Devaine et al., 2014)) and cooperative (Stag Hunt (Yoshida et al., 2008)) human behavior in strategic games. These models typically assume a fixed Level-0 strategy such as Win-Stay Lose-Shift ($WSLS$). Yet behavioral evidence suggests participants also employ strategies like $IMITATE$ or $Tit-for-Tat$ (Axelrod, 1980).

To address this, we introduce a new ToM model that estimates a dynamic belief over multiple strategies at Level-0 . This allows the model to generate more accurate predictions by dynamically inferring the current strategy. It also recursively builds higher ToM levels by computing best responses on the basis of these beliefs. In the following sections, we introduce the model and detail its structure, with particular emphasis on strategy inference at Level-0, given its critical role in shaping the ToM hierarchy. We then demonstrate the model's ability to accurately infer ToM depth. This work represents a first step toward a more comprehensive framework for modeling low-level strategy uses in social interactions.

Methods

To estimate opponent behavior at Level-0, the model evaluates observed actions against a set of strategies $S = s_1, ..., s_n$, such as $WSLS$, $IMITATE$, $REPEAT$ and $OPPOSE$. For each strategy s, the model maintains a belief weight α_s, updated using a retention factor ζ and a learning rate η (Smith et al., 2022):

$$\alpha_s \leftarrow \zeta \times \alpha_s + \varepsilon \times b(c_{OP}|s)$$

where $b_0(c_{op}|s)$ is a binary indicator over opponent actions, assigning 1 to the action that matches a strategy s, and 0 otherwise. For instance, the $IMITATE$ strategy predicts that $c_{OP}(t) = c_{AG}(t-1)$, so if $c_{AG}(t-1) = 1$, then $b_0(c_{OP}(t)|s = IMITATE) = [1,0]$. This defines a deterministic belief over the opponent's next choice under strategy s. The Dirichlet distribution over strategies is thus obtained as:

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

which is used to compute the integrated Level-0 belief over opponent choices as:

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

When it comes to recursive ToM reasoning, the model computes expected values using the utility matrix $U(c_{AG}, c_{OP})$, which depends on the choice of the agent c_{AG} and the choice of the opponent c_{OP}. Higher ToM levels are constructed as best responses to the predicted opponent's behavior at the level below. A level-k agent computes the expected value of each action by simulating an opponent at level $k-1$:

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

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

Importantly, only Level-0 beliefs are updated across trials; higher levels belief fall out of the recursive 'best-response' reasoning (De Weerd et al., 2018; de Weerd et al., 2013)

Results

We validated the model using simulated data from Matching Pennies (MP) under varying levels of volatility. Choice behavior was generated using an L1-agent. The model reliably recovers both learning and retention parameters across a grid of true values at two levels of volatility (Figure 2, top and middle row) (low: 20%, high: 80%). All recovered parameters exhibit a strong Pearson correlation with the true parameters

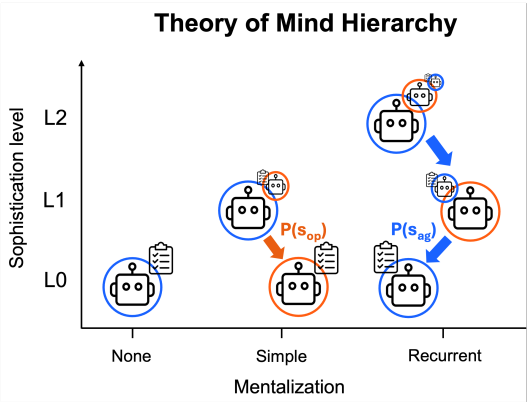

Figure 1: Hierarchy over ToM levels. At Level 0, the agent (AG) follows action strategies without considering the strategies used by the opponent (OP). At Level 1, the agent estimates a probability distribution $P(s_{OP})$ over opponents actions, marginalizing over the belief over strategies. At Level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

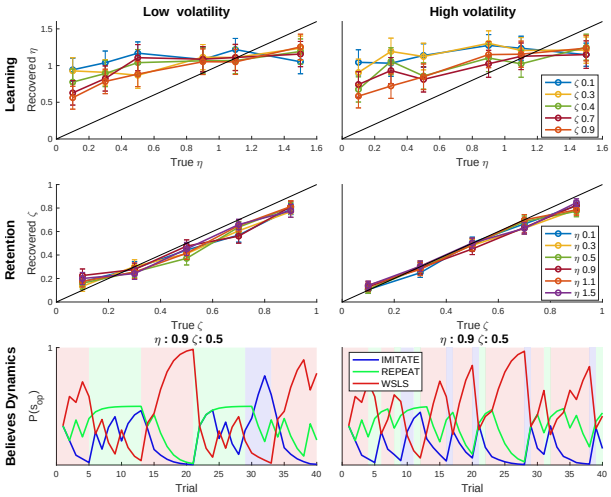

Figure 2: Model validation using simulated data from a volatile Matching Pennies (MP) environment. Top row: Parameter recovery plots for the learning rate across under low (left) and high (right) volatility conditions. Middle row: Parameter recovery plots for the retention rate. Each line represents a different fixed value of the complementary parameter; error bars indicate the standard error of the mean (SEM) across simulations. Bottom row: Trial-wise belief dynamics showing the model's inferred probabilities for each Level-0 strategy. Background color indicates the true strategy used by the simulated opponent. Note: In MP, $WSLS$ and $OPPOSE$ are behaviorally indistinguishable.

$(p < 0.001)$, with the lowest correlations observed at high volatility: $r > 0.72$ for learning and $r > 0.98$ for retention. The model also flexibly tracks and adapts its strategy beliefs in response to environmental changes (Figure 2, bottom row).

We also evaluated model's ability to identify the correct Theory of Mind (ToM) level. Specifically, we generated 1,000 behavioral sequences using either a Level-1 or a Level-2 agent. Each sequence was fit using the true parameters, and cumulative negative log-likelihoods were computed under both Level-1 and Level-2 models. The model reliably identified the correct ToM level, with a highly significant difference between the levels (real L1: one-side paired t-test: $t > 83$, $p < 0.001$; real L2: one-side paired t-test: $t < -88$, $p < 0.001$).

Discussion

This work introduces several key innovations to address limitations of previous ToM models: (i) mapping behavior onto interpretable strategies, (ii) accumulating evidence about strategies, (iii) adapting to shifts in strategy use, and (iv) computing higher-level decisions as best responses to inferred lower-level beliefs. In addition, uncertainty over ToM depth is modeled via a second Dirichlet distribution, with weights updated according to each level's predictive accuracy (not presented here).

Our results show that the model reliably recovers learning and retention parameters across varying conditions. Notably, recovery of learning rates is limited in certain regimes, which could be due to their secondary influence on action probabilities: learning primarily shapes the Dirichlet beliefs over strategies, while actions are generated by integrating over these beliefs. Moreover, although belief dynamics appears noisy, it is an expected outcome in matrix games, where behavioral overlap between strategies is common. Crucially, the model is designed to handle this ambiguity by integrating across multiple strategies, enabling stable and accurate inference.

Future work should explore model's generalizability to other matrix games (e.g. Bach or Stravinsky, Prisoner's Dilemma, Stag Hunt) and more complex interactive settings (e.g. Social Foraging Task).

Acknowledgments

We are grateful for helpful support and conversation with Oleg Solopchuk . This work was funded by the German Research Council DFG, Collaborative Research Center SFB 1528 "Cognition of Interaction". All authors declare no conflict of interest.

References

Axelrod, R. (1980). More effective choice in the prisoner's dilemma [Publisher: SAGE Publications Inc]. *Journal of Conflict Resolution*, *24*(3), 379–403.

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).

Devaine, M., Hollard, G., & Daunizeau, J. (2014). Theory of mind: Did evolution fool us? (T. Zalla, Ed.). *PLoS ONE*, *9*(2), e87619.

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.

Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, *107*, 102632.

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind (T. Behrens, Ed.). *PLoS Computational Biology*, *4*(12), e1000254.