Facilitating Human-AI Coordination through Computational Theory of Mind

Roderick Yang Terng Seow (yseow@andrew.cmu.edu)
Hoda Heidari (hheidari@cmu.edu)
Cleotilde Gonzalez (coty@cmu.edu)
Carnegie Mellon University, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Abstract

How can an AI teammate implicitly coordinate with a human? We address this question by integrating Instance-Based Learning (IBL), a cognitive theory of learning and decision making, with the level-k Theory of Mind framework. We hypothesize that coordination emerges when partners adopt complementary k-levels and when the higher k-level agent has an accurate model of their partner's cognitive processes. To test this hypothesis, we introduce a simultaneous-choice, multi-attribute task, where outcomes depend on interactions between choice features and agent decisions. Simulations of pairs of IBL-based agents at different k-levels support the hypothesis that complementary k-levels enhance coordination. However, empirical results from an experiment reveal no advantage of [human, IBL-L2] pairs over [human, IBL-L1] pairs, even when participants are restricted to operate as L1 agents. Post-hoc simulations show that model fitting recovers some advantage for [human, IBL-L2] teams by enabling the IBL-L2 agent to more accurately predict their human partner's actions.

Keywords: human-AI coordination; cognitive modeling; instance-based learning; ToM

Introduction

As autonomous agents become more ubiquitous, it is important that we understand how best to design AI agents for effective human-AI collaboration. In line with that goal, much work has been devoted to developing AI agents that are adaptive to a variety of situations and to human partners. Such methods include, but are not limited to, training a single agent on multiple alternative AI-generated policies (Hu et al., 2020) and creating a library of AI agents trained to coordinate with different human trajectories (Zhao et al., 2022).

These approaches to developing adaptive AI agents generally assume that human policies are static. However, by interacting with the environment and learning about the association between their actions and outcomes, humans are likely to adapt their strategies and actions (Hoffman et al., 2024). As a result, any static policy is likely to become incompatible as the human gains more experience with the environment. Thus, to coordinate well with humans, AI agents and partners need to be equipped with models that can predict how human behavior changes as a function of experience.

One promising direction is to leverage computational cognitive models of human learning. In a recent study, Liu et al., 2021 developed an AI task scheduler that integrated dynamic predictions of human task completion times derived from a well-known model of human competence as an exponential function of experience (Ritter & Schooler, 2001). They demonstrate that incorporating such a learning model in the AI scheduler greatly reduces the AI's error in estimating humans' task completion times, which in turn improves the AI's ability to appropriately assign tasks to individuals.

In this study, we adopt a similar approach to develop AI partners in the context of shared workspace tasks, where independent and fully autonomous agents must coordinate to achieve team goals. For example, during a disaster scenario, medics and firefighters need to coordinate where they are going to search and rescue injured and potential victims, while maximizing their efficiency and safety of the rescue effort. Barring communication between agents, an agent's ability to choose the optimal action depends on both their understanding of the environment and on their ability to anticipate and predict their partners' actions (Gulati et al., 2021).

Theory of Mind (ToM) has been suggested as a solution to achieve implicit coordination - the process of aligning actions without communication - between team members (Stowers et al., 2021). Of particular relevance is the class of simulation theories, which proposes that ToM involves adapting our own cognitive mechanism to simulate and predict the behavior of others (Gordon & Cruz, 2003; Gurney et al., 2021). To enable ToM in artificial agents, past research has used simulation-based techniques to facilitate inverse planning; by assuming that an agent (A) acts in a way that optimizes their goals, a partner agent (B) can leverage Bayesian inference or inverse reinforcement learning to infer A's goals and thus anticipate A's next action (Baker et al., 2009; Jara-Ettinger, 2019).

Instead of inverse planning, we propose to implement ToM in AI agents through simulating how human agents' knowledge and decisions change with experience. A promising candidate theory is Instance-Based Learning (IBL) - a theory of human learning and decision making that has successfully explained human behavior across a variety of decision-making contexts (Cranford et al., 2021; Gonzalez, 2022; Gonzalez et al., 2003). We additionally embed IBL models within the k-level reasoning framework - a theory of recursive reasoning in strategic games (Zhang et al., 2012) - to account for varying degrees to which the AI agent needs to anticipate or predict its partner.

Based on simulations between pairs of IBL agents at different k-levels, we hypothesized that human-AI teams with complementary k-levels will outperform teams with identical k-levels, especially in situations that demand greater co-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

ordination. We then tested this hypothesis in an empirical experiment in which human participants were partnered with different AI agents. While the results of the experiment did not directly support the hypothesis, a follow-up simulation experiment showed that improving the model's prediction of human behavior through parameter fitting recovered some advantage for complementary k-level teams.

Background

Instance-Based Learning

Instance-Based Learning (IBL) is a theory of decision making based on experience that has successfully accounted for human decisions in a variety of tasks (e.g., Gonzalez, 2022; Gonzalez et al., 2003; Nguyen et al., 2023). The theory posits that experiences are represented as instances; each instance is a unit that binds contextual features s, the selected action a, and the experienced outcome or expected utility in the case where no outcome was observed u. When making a decision, the expected value of each option (called Blended Value in IBL theory) is computed as a weighted average of the outcome values u_i associated with previous experiences:

$$V(a) = \sum_i P_i u_i \tag{1}$$

The weight of each instance P_i is calculated as a softmax transformation over its memory activation A_i:

$$P_i = \frac{e^{A_i/\tau}}{\sum_j e^{A_j/\tau}} \tag{2}$$

where the activation of an instance depends on the addition of three components:

$$A_i = ln(\sum_j (t - t'_{i,j})^{-d}) + \mu \sum_k (w_k(Sim(s_{i,k}, s_{t,k}) - 1) + \sigma\xi \tag{3}$$

The first component captures the frequency and recency of the considered instance; due to the decay parameter d instances that are further back in time will have a lower activation. The second component captures similarity where μ determines the balance between similarity and recency on an instance's activation, w_k is the importance of a context feature for calculating similarity, and Sim is a specified function that returns the degree of similarity for a particular feature k between instance i ($s_{i,k}$) and the current context ($s_{t,k}$). The third component introduces some randomness to the memory retrieval process where $\sigma\xi$ is a scaled noise distribution.

K-level reasoning

K-level reasoning is a non-equilibrium theory of player decisions in adversarial games (Chong et al., 2016; Stahl & Wilson, 1995). The theory posits that players select the best response but vary in terms of the recursive depth involved in anticipating the actions of other players. Specifically, a level-k player responds assuming the other players are level-$k-1$ such that the base player (level-0) is assumed to act randomly, the level-1 player selects the best action assuming that other players are level-0, the level-2 player selects the best action assuming that other players are level-1, etc. Thus, obtaining the best payoff relies on accurately determining the k-level of your opponent(s), successfully predicting their actions, and then selecting the action that would result in the highest payoff conditional on the predicted actions of other players.

A similar principle can be applied to the case of human-AI coordination, particularly in situations where the team reward relies on both the independent actions of the team members and the interaction between those actions. We hypothesize that a team with members of complementary k-levels would perform better than a team with members of identical k-levels. In the complementary team, the actions of level-n agents would be accurately predicted by the level-n+1 agents and the level-n+1 agents would be able to select the optimal action conditioned on its prediction of its level-n partner.

To our knowledge, there have been no prior attempts to model purely collaborative settings with the k-level theory. Furthermore, k-level has been applied only to scenarios in which the optimal decision for each k remains static. This stands in contrast to many multi-agent collaborative settings, where team members have to simultaneously learn how their joint actions affect the team payoff.

Methods

Victim Rescue Task

The Victim Rescue Task is a two-player collaborative multi-attribute decision-making task that combines learning and game-theoretic properties. On the learning side, this task is similar to more abstract function learning tasks, where a regular function underlies the relationship between option attributes and the payoff (e.g., Koh and Meyer, 1991). It also shares characteristics with contextual bandit tasks, where the agent has to choose between a number of options, and the payoff of each option depends on the features of each option or on some common context (e.g., Niv, 2009).

On the game-theoretic side, this task extends the matrix games commonly used to study decision making. However, unlike typical two-player games, the payoffs associated with each action are neither fixed nor revealed beforehand, and thus players have to learn how their joint actions predict payoffs through repeated interactions with the task environment.

The overall goal of the team (a medic and a firefighter) is to maximize the number of rescued victims from a burning building. In each trial, players decide which of two rooms to visit, where each room is defined by three attributes: distance to the building entrance, room size, and intensity of the fire in the room. Players assume the role of a medic or firefighter, where each role has slightly different abilities that interact with room attributes to affect the team payoff.

The number of victims available in a room (x) is:

$$Ava(x) = 150/(1 + exp(-0.01 * (size - 400))) \tag{4}$$

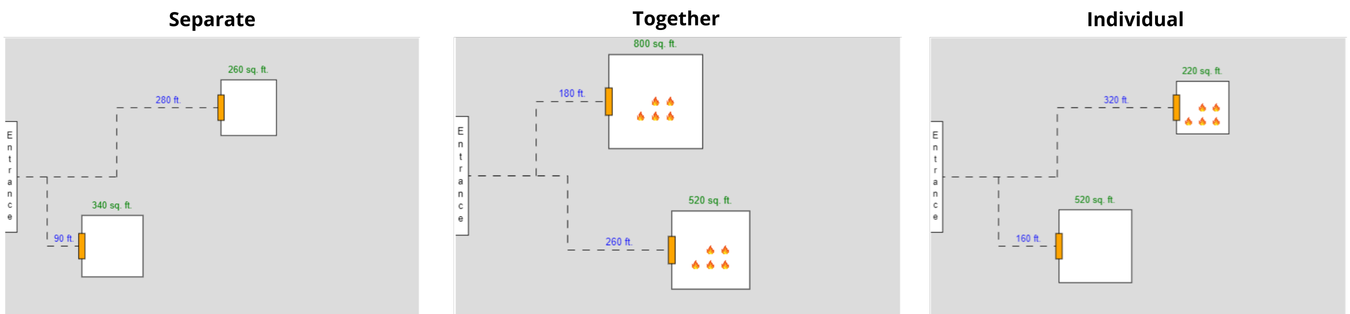

Figure 1: The three scenarios (trial types). Let $R(x,y)$ represent the team payoff resulting from the medic choosing room x and the firefighter choosing room y. For "Separate" trials: $R(1,2) <> R(2,1) > R(1,1) <> R(2,2)$. For "Together" trials: $R(1,1) <> R(2,2) > R(1,2) <> R(2,1)$. For "Individual" trials: $R(1,1) > R(1,2) <> R(2,1) > R(2,2)$ or $R(2,2) > R(1,2) <> R(2,1) > R(1,1)$.

The maximum number of victims that can be rescued by each role is:

$$Abi(x,f) = 40 - 1 * fire - 0.05 * distance \tag{5}$$

$$Abi(x,m) = 70 - 5 * fire - 0.1 * distance \tag{6}$$

If the firefighter and medic both choose the same room, then the firefighter is able to help the medic out by suppressing the fire, which changes the medic's fire coefficient to 1:

$$Abi(x,m) = 70 - 1 * fire - 0.1 * distance \tag{7}$$

If the two agents choose different rooms, then the total number of victims rescued is the sum of their ability scores:

$$Reward = Abi(x_m, m) + Abi(x_f, f) \tag{8}$$

If the two agents choose the same room, then it is possible that the combined abilities of the agents exceed the number of victims in the chosen room:

$$Reward = min(Ava(x), Abi(x,m) + Abi(x,f)) \tag{9}$$

Based on this payoff structure, we designed three types of coordination scenarios. On "individual" trials, one room is always dominant; that is, the optimal action for either agent is to select that room regardless of the other's choice. This reflects a situation where one room has many more victims than the other room, and thus rescuers should focus their efforts on that room regardless of where their partner goes. In "together" trials, the optimal team-level joint action is for the two agents to choose the same room. This reflects a situation where both rooms contain many victims and high fire intensities, and therefore rescuers need to choose the same room, allowing the firefighter to suppress the fire in the room and increasing the rescue efficiency of the medic. In "separate" trials, the optimal team-level joint action is for the two agents to choose different rooms. This reflects a situation where both rooms contain a small number of victims, and thus rescuers need to split their efforts in order to maximize the amount of ground covered.

30 of each type of trial were randomly generated for a total of 90 trials, and 10 sequences were generated from the same set of 90 trials with the constraint that each trial type would be encountered once before the same trial type was encountered again. In other words, the trials were organized into blocks of 3 for a total of 30 blocks, where each block would contain one trial of each type.

Partner agents

We developed two types of cognitive agents and two reference agents. The cognitive agents are illustrated in Fig. 2. The first type of cognitive agent is the level-1 IBL agent (IBL-L1) which learns to associate the payoff from the environment with their own actions and ignore the actions of the other. When deciding which room to select, the IBL-L1 agent relies on the IBL mechanisms and selects the room with the highest expected utility based on the features that define each room. Upon receiving feedback, the IBL-L1 agent creates and stores an instance that represents its own action and the observed outcome.

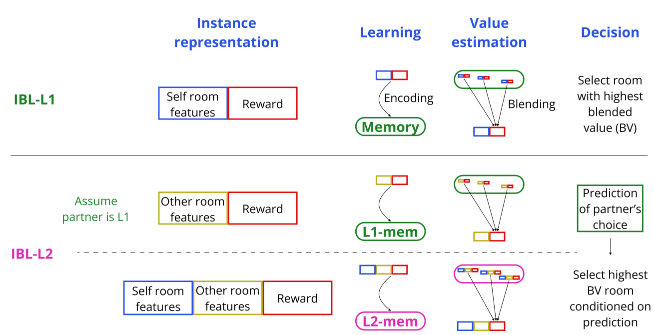

Figure 2: Schematic of proposed IBL agents.

In contrast, a level-2 IBL agent (IBL-L2) learns to associate the payoff from the environment with both their own and their partners' actions. During room selection, the IBL-L2 agent assumes that its partner is an IBL-L1 agent and correspondingly makes a prediction of which room its partner will choose. The IBL-L2 agent then selects what it estimates to be the best room, conditioned on the prediction of its partner's chosen room. Upon receiving feedback, the IBL-L2 agent forms an instance that represents the features of its chosen room, the features of its partner's chosen room, and the observed outcome. The IBL-L2 then aligns its internal IBL-L1 model with its partner through model-tracing (Anderson et al., 1995) by creating an instance that reflects the features of its partner's actual chosen room and the observed outcome.

For both types of IBL agents, we used the default IBL parameter values: decay (0.5), noise (0.25), temperature ($\sqrt{2} * 0.25$), attribute weights (1 for room size, distance, and fire intensity).

Simulation-Based Predictions

Before running the experimental study with human participants, we simulated the performance of pairs of agents on ten different trial sequences. There were four teams; two teams with complementary k-levels (IBL-L1 + IBL-L2 and IBL-L2 + IBL-L1) and two teams with identical k-levels (IBL-L1 + IBL-L1 and IBL-L2 + IBL-L2).

The simulation results suggest that teams with agents of complementary k-levels demonstrate the best performance, particularly on separate-type trials, whereas agents with identical k-levels perform poorly on separate-type trials (refer to Figure 3. These results are consistent with the hypothesized advantage for teams with complementary k-level agents because the higher k-level partner is able to accurately predict its lower k-level partner's actions, and thus select the optimal team-level action.

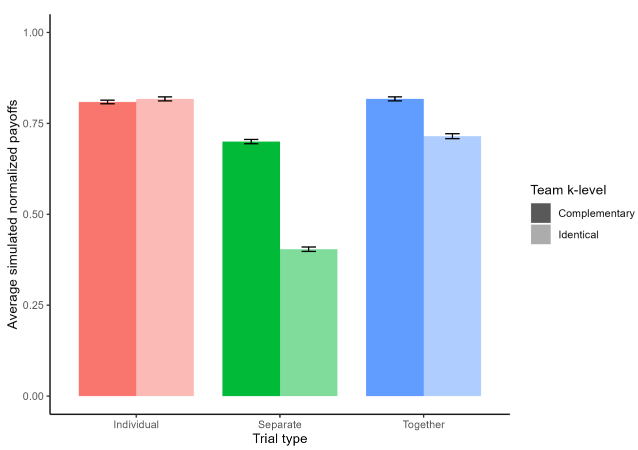

Figure 3: Simulated average performance between pairs of AI agents in the Individual, Separate and Together tirals. Notably, teams with complementary k-levels (i.e., IBL-L1 + IBL-L2 or IBL-L2 + IBL-L1) show a clear performance advantage on Separate trials and to a lesser extent in the Together trials, compared to teams with identical k-levels (i.e., IBL-L1 + IBL-L1 or IBL-L2 + IBL-L2). This advantage of complementary k-levels is not observed in Individual trials.

Experimental procedure

240 participants were recruited online from Amazon Mechanical Turk. Each participant was randomly assigned a role (medic or firefighter), paired with one of the four types of possible AI partners, and assigned to one of the ten trial sequences. Participants were compensated with a base pay of \$3 and a performance bonus of up to \$6 depending on their performance in "individual" type trials. Demographic data was not collected for this study.

Participants encountered four attention trials throughout the experiment, where instead of selecting the room that would provide the largest payoff, participants were explicitly instructed to select the room that fulfilled a given condition (e.g., select the smaller room). Participants who failed more than one attention check were excluded from the analysis. We also excluded participants who submitted irrelevant or nonsensical responses to open-ended questions in the post-experiment survey. 194 participants remained in the sample after exclusion.

As this is a preliminary study, we wanted to verify that our proposed IBL-L1 agent would be an appropriate model of human choices in the Victim Rescue task. Thus, participants were constrained to operate at best as an L1 agent by hiding all information about their partners' choices, essentially reducing the task to a single-player task from the participant's perspective.

Empirical Results

Figure 4 displays the average payoff per trial for all participants in each condition (partner type) and trial type. To make payoffs between trials comparable, payoffs were normalized according to the maximum and minimum payoffs available for each trial.

As expected, teams with an optimal partner outperformed all other teams, with an average normalized payoff of 0.86. Teams with a random agent performed close to chance (0.54). Teams with an IBL-L2 partner (0.64) performed comparably to teams with an IBL-L1 partner (0.65). To jointly estimate the effect of partner type, trial type, and experience (block), a linear regression model with these factors as independent variables and the normalized payoff per trial was fit. For ease of interpretation, the reference partner type was IBL-L1, and the reference trial type was "together". An ablation analysis comparing models with various combinations of predictor variables found that the full regression model best predicted the outcome variable (Adjusted $R^2 = 0.133$).

As the regression results in Table show, teams generally improved with experience ($\beta = 0.002$, $SE = 0.0003$, $p < 0.01$). Consistent with simulation results, teams also performed better in individual trials compared to "together" trials ($\beta = 0.125$, $SE = 0.006$, $p < 0.01$).

However, we did not find a statistically significant advantage for IBL-L2 compared to teams with partners of IBL-L1 ($\beta = -0.009$, $SE = 0.008$, $p = 0.246$). Unlike simulated performance, there was also no reliable advantage for the "to-

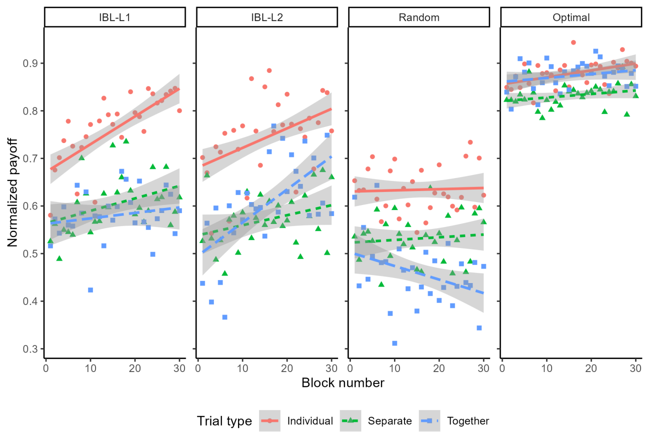

Figure 4: Average normalized payoff across participants by partner type and trial type.

	Beta	S.E.	f^2
Block	0.002^{***}	0.003	0.0020
Bot: L2	-0.009	0.008	0.124
Bot: Rand.	-0.107^{***}	0.007	
Bot: Opt.	0.212^{***}	0.007	
Trial: Sep.	0.009	0.006	0.0273
Trial: Ind.	0.125^{***}	0.006	
Intercept	0.577^{***}	0.008	
Observations	17,460		
R^2	0.133		
Adjusted R^2	0.133		
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 1: Coefficients from full regression model predicting normalized payoff per trial. f^2 values report the effect size for each variable (block, partner type (bot), and trial type).

gether" trials compared to the "separate" trials ($\beta = 0.009$, $SE = 0.006$, $p = 0.165$).

Model Fitting

One possible explanation for the lack of empirical advantage observed in the team performance of [human, IBL-L2] teams is that the IBL-L2 agent cannot accurately predict the choices of its human partners. We can quantify this prediction accuracy by calculating the number of matches between each model's predictions and each participant's choices. This quantity will be termed ”“model sync”. Because the IBL-L2 agent models its partner using an IBL-L1 model with fixed default parameter values, we explored whether allowing model parameters to vary through model fitting would improve model sync between IBL-L1 models and humans. Furthermore, we tested whether integrating fitted IBL-L1 models into IBL-L2 agents would improve team performance in a simulation experiment.

Each participant's choices (from all conditions) were fit using an IBL-L1 model. Similarly to how the IBL-L2 agent uses model-tracing to align its internal IBL-L1 model with its partner, model-tracing was also in effect during fitting. Model sync was optimized per participant by applying a differential evolution algorithm to search the space of possible parameter values (Virtanen et al., 2020). Note that for model fitting, the noise parameter was constrained to 0 to allow for stable parameter convergence, which yielded 5 free parameters per fit.

Across all conditions, the IBL-L1 model with default parameters only achieved a maximum synchronization with human actions of 65% at the end of the experiment. This means that an IBL-L2 agent incorrectly predicted the choice of its human partner a third of the time and thus was unable to correctly complement its partner despite having learned the payoff structure of the task. In contrast, the IBL-L1 model with parameters fitted to individuals achieved a model sync of 80% by block 5, which would have allowed the IBL-L2 agent to coordinate faster and more accurately with its human partner. A logistic regression predicting model sync in each trial confirms that model sync improved significantly with model fitting ($\beta = 0.793$, $SE = 0.024$, $p < 0.01$).

Consistent with the lack of an observed difference in team performance, post-hoc explorations revealed no significant differences between conditions in terms of the distributions of fitted parameter values. In particular, between participants who were paired with an IBL-L1 agent and those paired with an IBL-L2 agent, Wilcoxon rank-sum tests indicated that there were no significant differences between the decay values ($W = 997$, $p = 0.66$) and temperatures ($W = 926$, $p = 0.88$), nor the attribute weight values of the room size ($W = 1167$, $p = 0.06$), room distance ($W = 1026$, $p = 0.50$), and fire intensity ($W = 1049$, $p = 0.38$).

Simulation Experiment

A challenge for future work is to determine the values of individual human cognitive parameters before interacting with AI, so that the predictions of individual actions by IBL-L1 can be more accurate. To test this idea, we performed a simulated experiment. The experiment explores whether integrating these fitted IBL-L1 models into IBL-L2 agents may consequently improve team performance.

We took the sequence of choices for each participant that was partnered with an IBL-L1 or IBL-L2 agent and simulated the task with the default and individually fitted parameter versions of the IBL-L1 model. For those originally partnered with an IBL-L1 agent, they would instead be paired with a fitted IBL-L1 model. For those partnered with an IBL-L2 agent, they would instead be paired with an IBL-L2 agent that contained a fitted IBL-L1 model. It is important to note that unlike a real-time experiment, the "human" partner in these simulations is a predetermined sequence of choices and thus does not reflect how an actual human partner might respond to changes in feedback that result from the fitted models making potentially different decisions on each trial as compared to the default models.

Given that model fitting improves the alignment between IBL-L1 models and human participants, we hypothesized that IBL-L2 agents with a fitted IBL-L1 model would better predict their human partner and consequently improve team performance, particularly in "separate" trials. In contrast, IBL-L1 agents based on fitted models were hypothesized to better mimic the behavior of their human partner, which would decrease the team performance in "separate" trials.

A linear regression predicting payoff per trial was fitted as a dependent variable and the independent variables of interest were block number, partner type, trial type and model fit. To test our hypothesis, the regression model included interactions between all independent variables.

Aligned with our hypothesis, the four-way interaction (block*L2*fitted*separate) confirms that model fitting improves team performance, but only for teams with IBL-L2 agents in "separate" trials ($\beta = 0.00841$, $SE = 0.00342$, $p < 0.05$). As the graphs in Figure 5 for "separate" trials show, the performance advantage for IBL-L2 agents with fitted models over their default counterparts increases with block number, whereas the opposite is true for IBL-L1 agents.

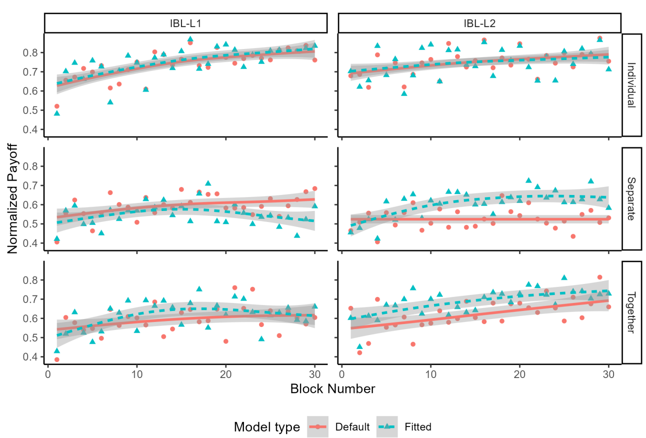

Figure 5: Simulated team performance on experiment with individually fitted IBL-L1 models, either directly as an IBL-L1 agent or integrated into an IBL-L2 agent.

Discussion and Conclusions

Without the ability to explicitly communicate and establish a shared mental model, team members must rely on ToM to accurately anticipate the actions of their teammates to achieve effective coordination (Gulati et al., 2021). This study explored how to develop AI agents for effective human-AI team coordination by implementing simulation-based ToM in AI agents (Gurney et al., 2021). In contrast to previous research on human-AI teaming, which has mainly assumed static models of human behavior (Hoffman et al., 2024), we modeled the dynamics of how decision-making processes change with experience using a cognitive model of human learning and decision making, Instance-Based Learning (Gonzalez et al., 2003).

According to the k-level reasoning framework (Zhang et al., 2012), the most effective pair would consist of agents at complementary k-levels, such that the level-n+1 partner would predict the actions of the level-n partner. Simulations between pairs of IBL agents predicted that teams with complementary k-levels achieved the highest team performance. However, empirical experiments with human-AI teams revealed that there was no significant advantage for teams that contained an IBL-L2 compared to those with an IBL-L1 even when human agents were restricted to operate as L1, since we hid any information about the partner.

One possible explanation for this discrepancy between the predictions and empirical results is that the default IBL model is unable to predict the actions of human agents with high accuracy, which in turn may impair the IBL-L2 agent's ability to select actions that optimize team-level outcomes. Consistent with the idea that optimal team-level action selection depends on accurately predicting other team members' actions, further simulations using the collected human data confirmed that IBL-L2 agents with fitted IBL-L1 models outperform their non-fitted counterparts on "separate" trials - situations where coordination relies on accurately predicting one's partner's actions. Thus, an exciting direction for future research is to develop methods for efficient online model fitting, which would allow the IBL-L2 agent to flexibly adapt to any human partner and to coordinate better with them.

Another explanation for the discrepancy is the human potential ability to operate at a higher level than L1, despite the lack of information about the partner. This hypothesis is supported by empirical evidence from neuroscience for the idea that internal simulation mechanisms may contribute to our ability to understand and predict the mental states of others (Gallese & Goldman, 1998). That is, humans may have the capability to use self-projection or their own learning from the rewards observed with repeated experience to predict the mental states of others. In many real-world team tasks, the emergence of ToM in humans may likely rely on indirect, generalized, or internally generated sources of knowledge to construct mental state attributions. Future research should investigate how an AI partner could learn whether its human partner was engaging in ToM and proactively adapting to AI actions.

Acknowledgments

This material is based upon work supported by the AI Research Institutes Program funded by the National Science Foundation under AI Institute for Societal Decision Making (AI-SDM), Award No. 2229881.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The journal of the learning sciences*, *4*(2), 167–207.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Chong, J.-K., Ho, T.-H., & Camerer, C. (2016). A generalized cognitive hierarchy model of games. *Games and Economic Behavior*, *99*, 257–274.

Cranford, E. A., Gonzalez, C., Aggarwal, P., Tambe, M., Cooney, S., & Lebiere, C. (2021). Towards a cognitive theory of cyber deception. *Cognitive Science*, *45*(7), e13013.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493–501.

Gonzalez, C. (2022). Learning and dynamic decision making. *Topics in Cognitive Science*, *14*(1), 14–30.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Gordon, R. M., & Cruz, J. (2003). Simulation theory. *Encyclopedia of cognitive science*, *4*, 9–14.

Gulati, A., Nguyen, T. N., & Gonzalez, C. (2021). Task complexity and performance in individuals and groups without communication. *AAAI Fall Symposium*, 102–117.

Gurney, N., Marsella, S., Ustun, V., & Pynadath, D. V. (2021). Operationalizing theories of theory of mind: A survey. *AAAI Fall Symposium*, 3–20.

Hoffman, G., Bhattacharjee, T., & Nikolaidis, S. (2024). Inferring human intent and predicting human action in human–robot collaboration. *Annual Review of Control, Robotics, and Autonomous Systems*, *7*.

Hu, H., Lerer, A., Peysakhovich, A., & Foerster, J. (2020). "other-play" for zero-shot coordination. *International Conference on Machine Learning*, 4399–4410.

Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, *29*, 105–110.

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 811.

Liu, S., Yao, S., Fu, X., Shao, H., Tabish, R., Yu, S., Bansal, A., Yun, H., Sha, L., & Abdelzaher, T. (2021). Real-time task scheduling for machine perception in intelligent cyber-physical systems. *IEEE Transactions on Computers*, *71*(8), 1770–1783.

Nguyen, T. N., Phan, D. N., & Gonzalez, C. (2023). Speedyibl: A comprehensive, precise, and fast implementation of instance-based learning theory. *Behavior Research Methods*, *55*(4), 1734–1757.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. *International encyclopedia of the social and behavioral sciences*, *13*, 8602–8605.

Stahl, D. O., & Wilson, P. W. (1995). On players models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*(1), 218–254.

Stowers, K., Brady, L. L., MacLellan, C., Wohleber, R., & Salas, E. (2021). Improving teamwork competencies in human-machine teams: Perspectives from team science. *Frontiers in Psychology*, *12*, 590290.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272.

Zhang, J., Hedden, T., & Chia, A. (2012). Perspective-taking and depth of theory-of-mind reasoning in sequential-move games. *Cognitive science*, *36*(3), 560–573.

Zhao, M., Simmons, R., & Admoni, H. (2022). Coordination with humans via strategy matching. *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 9116–9123.