Modeling Deep Strategic Reasoning by Humans in Competitive Games

Xia Qu
Dept. of Computer Science
University of Georgia
Athens, GA 30602
quxia@uga.edu

Prashant Doshi
Dept. of Computer Science
University of Georgia
Athens, GA 30602
pdoshi@cs.uga.edu

Adam Goodie
Dept. of Psychology
University of Georgia
Athens, GA 30602
goodie@uga.edu

Abstract

The prior literature on strategic reasoning by humans of the sort, *what do you think that I think that you think*, is that humans generally do not reason beyond a single level. They think about others' strategies but not about others' reasoning about their strategies. When repeatedly faced with another who reasons about their strategies, humans learn to think one level deeper but the learning is generally slow and incomplete. However, recent evidence suggests that if the games are made competitive and therefore representationally simpler, humans generally exhibited behavior that was more consistent with deeper levels of recursive reasoning. We seek to computationally model the judgment and behavioral data that is consistent with deep recursive reasoning in competitive games. We present process models built from agent frameworks that not only simulate the observed data well but also exhibit psychological intuition. Our goal is to gain insights into the strategic thinking process, ultimately leading to agents which can emulate human decision making and effectively interact with humans in mixed settings.

Introduction

An important aspect of strategic reasoning is the depth to which one thinks about others' thinking about others' in order to decide on an action. Investigations in the context of human recursive reasoning (Ficici and Pfeffer 2008; Hedden and Zhang 2002; Stahl and Wilson 1995) reveal a pessimistic outlook: in general-sum games, humans think about others' strategies but generally do not ascribe further recursive thinking to others. Consequently, the default level of recursive reasoning tends to be shallower than imagined. When humans repeatedly experience games where others do reason about others' actions in deciding their own, humans learn to reason about others' reasoning, but the learning tends to be slow, requiring many experiences, and incomplete, a significant population continues to exhibit shallow thinking. [1]

[1]Note that these systematic results pertain to the general, adult population and should not be confused with anecdotal evidence of deeper thinking.

Recently though, Goodie et al. (2010) reported that in fixed-sum, *competitive* games human behavior is generally consistent with deeper levels of recursive reasoning. In games designed to test two and three levels of recursive reasoning, the observed actions were broadly consistent with these deeper levels by default. On experiencing these games repeatedly, the proportion of participants exhibiting the deeper thinking leading to rational behavior in those games increased even further, which is indicative of learning. In the context of the previous pessimism, these results are important in that they better align systematic observations with our expectations. Many of these experiments utilized a modification of Rosenthal's Centipede game (Rosenthal 1981), which is a two-player game well suited to evaluate how deep players think that the other could be reasoning about other's action.

Given the availability of these behavioral data, computationally modeling them may facilitate an understanding of the underlying cognitive processes. The models, if successful, may also provide insights toward ultimately building agents that emulate human strategic decision making and that effectively interact with humans in mixed settings. Frameworks for decision making in multiagent settings, and specifically those that integrate recursive reasoning, offer a suitable point of departure as potential models. One such framework is the interactive partially observable Markov decision process (I-POMDP) (Gmytrasiewicz and Doshi 2005), that generalizes the well-known POMDP to multiagent settings. An I-POMDP is particularly appropriate because it elegantly integrates modeling others and others' modeling of others in the subject agent's decision-making process. Previously, Doshi et al. (2010) utilized an empirically informed I-POMDP, simplified and augmented with psychologically plausible learning and choice models, to computationally model behavioral data pertaining to recursive reasoning up to the *second level*. Data from both general- and fixed-sum games, providing evidence of predominantly level 1 and level 2 reasoning, respectively, was successfully modeled.

In this paper, we model human judgment and behavioral data, reported by Goodie et al. (2010), that

is consistent with *three* levels of recursive reasoning in the context of Centipede games. In doing so, we investigate principled modeling of data up to levels rarely performed before. The previous I-POMDP based model utilized underweighted belief learning, parameterized by γ, and a quantal response choice model (McKelvey and Palfrey 1995) for the subject agent parameterized by λ. We extend this model to make it applicable to games evaluating up to level three reasoning. Although it employs an empirically supported choice model for the subject agent, it does not ascribe plausible choice models to the opponent who in the experiments is also projected as being human. In previous work, the performance of this modeling on prediction scores is not as good as that on achievement scores. The simulated prediction scores using this model are higher than that in the study data. Hence, our second candidate model generalizes the previous by intuitively utilizing a quantal response choice model for selecting the opponent's actions at level 2. We evaluate how well these models fit the data and simulate it, and also compare between the two. The competitive nature of the game discourages the influence of essentially cooperative social constructs such as positive reciprocity and altruism, otherwise prevalent in strategic games (Camerer 2003). However, other processes such as inequality aversion may not be ruled out, and we analyze its relevance.

Background: Level 3 Recursive Reasoning

The experiments utilized a two-player alternate-move game of complete and perfect information. In order to test level 3 recursive reasoning, Goodie et al. (2010) extended the Centipede game to five states. The game and its tree are depicted in Fig. 1. It starts at state A where player I may choose to *move* or *stay*. If player I chooses to move, the game goes to state B where player II needs to decide between moving or staying. If a move is taken, the game proceeds to the next state and the other player takes its turn to choose. Game continues up to two moves of player II. An action of stay by either player also terminates the game. In the study, the focus is on how player I plays the game when it starts at A.

Outcomes on staying or when the game terminates at E are probabilities of winning in a different task for each player. In Fig. 1(b), the outcomes are for player I, player II's outcomes are one minus player I's outcomes. Rational choice is the action that maximizes the probability of winning. In order to decide whether to move or stay at state A, a rational player I must reason about whether player II will choose to move or stay at B. A rational player II's choice in turn depends on whether player I will move or stay at C. And a rational player I's choice in turn depends on whether player II will move or stay at D. Thus, the game lends itself naturally to recursive reasoning and the level of reasoning is governed by the height of the game tree.

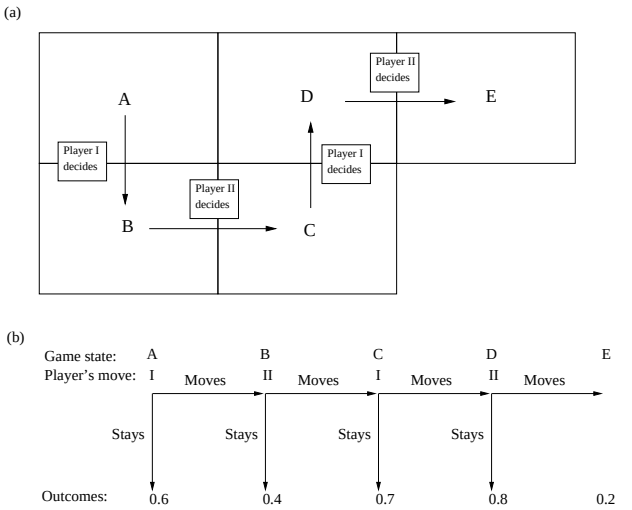


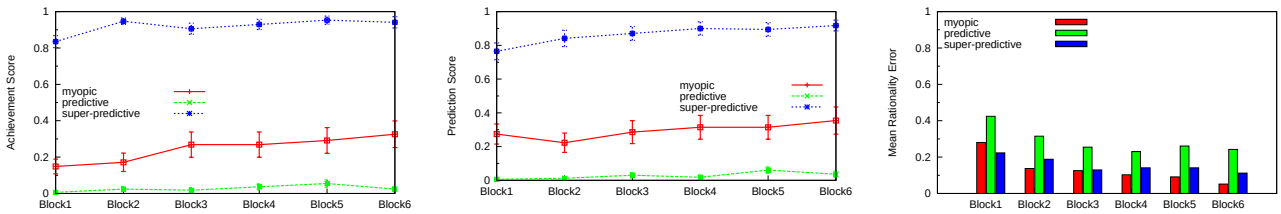
Figure 1: A four-stage, fixed-sum, sequential game in (a) matrix and (b) tree form. The fixed-sum payoffs are different from those in Rosenthal's Centipede game.

Methodology

In order to test up to the third level of recurse reasoning, Goodie et al. designed the computer opponent (player II), projected as a human player, to play a game in three ways if player I chooses to move and game proceeds to state B: (i) Player II decides on its action by simply choosing rationally between the default outcomes of staying at states B and C in Fig 1(b); II is a zero-level player and is called *myopic*. (ii) II reasons that player I will choose rationally between the default outcomes of stay at C and stay at D, and based on this action, II selects an action that maximizes its outcomes; II is a first-level player who explicitly reasons about I's subsequent choice and is called *predictive*. (iii) II reasons that player I is predictive and who will act rationally at C reasoning about II's rational action at D; II is a second-level player who explicitly reasons about I's subsequent choice which is decided by rationally thinking about II's subsequent action at D. We call this type of II as *super-predictive*. Therefore, if player I thinks that II is super-predictive, then I is reasoning deeply to three levels.

To illustrate, in the game of Fig. 1(b), if player I chooses to move, then she thinks that a myopic player II will stay to obtain a payoff of 0.6 at state B compared to move which will obtain 0.3 at state C. She thinks that a predictive II thinking that I being myopic will move to D thereby obtaining 0.8 instead of staying at C, will decide to move thinking that he can later move from D to E, which gives II an outcome of 0.8. A super-predictive II knows that I is predictive, knowing that if I moves from C to D then II will move to E which gives I only 0.2, hence I will choose to stay at C, therefore II will stay at B.

The rational choice of players in the game depends on the preferential ordering of states of the game rather



(a) (b) (c)

Figure 2: (a) Mean achievement scores (b) mean prediction scores and (c) mean rationality errors of participants for the opponent groups across test blocks.

than specific probabilities. Let $a \prec b$ indicate that the player prefers state b over a. In contrast to three-stage sequential games, we cannot find a single preference ordering in four-stage games that distinguishes between the actions of the three opponent reasoning types. Therefore, Goodie et al. employed two differently ordered games to diagnose each level of reasoning from the other two.

The game depicted in Fig. 1(b), which has the preference ordering of $\mathbf{E} \prec \mathbf{B} \prec \mathbf{A} \prec \mathbf{C} \prec \mathbf{D}$ for player I, is the only ordering that permits I's second level reasoning to be distinguished behaviorally from her third and first levels of reasoning. As we analyzed above, player I with level 2 reasoning will choose to move while with first and third levels, she will choose to stay. Preference orderings $\mathbf{C} \prec \mathbf{B} \prec \mathbf{A} \prec \mathbf{D} \prec \mathbf{E}$ and $\mathbf{C} \prec \mathbf{B} \prec \mathbf{A} \prec \mathbf{E} \prec \mathbf{D}$ are the two orderings which distinguish level 1 reasoning from levels 2 and 3. In these games, player I with level 1 reasoning will choose to move while with levels 2 and 3 will choose to stay.

Results

Participants were assigned randomly to different groups that played against a myopic, predictive or super-predictive opponent. Each participant experienced 30 trials with each trial consisting of two games whose payoff orderings are diagnostic and a catch trial controlling for inattention. For convenience of presentation, the 30 trials are grouped into 6 blocks of 5 trials each.

From the participants' data in each of the three types of opponent groups, Goodie et al. measured the *achievement score*, which is the proportion of trials in a block in which the participants played the conditionally rational action given the opponent's level of reasoning. They also reported the *prediction score*, which is the proportion of trials in which the participants' prediction about the opponent's action was correct given the opponent type, and the *rationality error* which measured the proportion of trials in which participants' actions were not consistent with their predictions of the opponent's actions.

In Figure 2(a), we show the mean achievement scores across all participants for each of the 3 opponent groups. [2] We define a metric, L, as the trial after which performance over the most recent 10 trials never failed to achieve statistical significance (cumulative binomial probability $< .05$). This implies making no more than one incorrect choice in any window of 10 trials. For participants who never permanently achieved statistical significance, L was assigned a value of 30. We observe that participants in super-predictive opponent group had the highest overall achievement score with an average L score of 11.3, followed by those in myopic opponent group with L score, 27.2. These L scores are consistent with the observations that achievement scores in these two groups are increasing. Participants in predictive opponent group had achievement score close to zero and L score of 30, which means they never truly achieved a corresponding strategy. Similar to the mean achievement score, participants in the super-predictive opponent group exhibited the highest prediction scores while those in the predictive opponent group had the lowest scores (Fig. 2(b)).

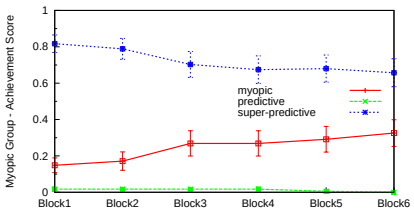


Figure 3: Mean achievement scores for just the myopic opponent group computed given myopic, predictive and super-predictive opponent types.

In order to investigate the comparatively low achievement scores for the myopic opponent group, we further computed the mean achievement scores for just this group given the three types of opponents, in Fig. 3. Notice that the achievement scores of this group given a super-predictive opponent are high indicating that a large proportion of participants are reasoning at the deepest level of 3, but this proportion gradually reduces due to learning.

Computational Modeling

We seek process-oriented and principled computational models whose predictions are consistent with the observed data of the previous section. These models dif-

[2] Our charts recomputed from the data may be slightly different from those in (Goodie, Doshi, and Young 2010).

fer from statistical curve fitting by providing some insights into the cognitive processes of judgment and decision making that possibly led to the observed data. In order to computationally model the results, a multiagent decision-making framework that integrates recursive reasoning in the decision process is needed. Finitely-nested I-POMDPs (Gmytrasiewicz and Doshi 2005) are a natural choice which meets the requirements of explicit consideration of recursive beliefs and decision making based on such beliefs.

Empirically Informed I-POMDP

Interactive POMDPs generalize POMDPs to multiagent settings by including other agents' models as part of the state space. A finitely-nested I-POMDP of agent i with a strategy level l interacting with another agent, j, is defined as the tuple:

$$\text{I-POMDP}_{i,l} = \langle IS_{i,l}, A, \Omega_i, T_i, O_i, R_i, OC_i \rangle$$

where:

- $IS_{i,l}$ denotes a set of interactive states defined as, $IS_{i,0} = S$ and $IS_{i,l} = S \times M_{j,l-1}$ for $l \geq 1$, where S is the set of states of the physical environment; $M_{j,l-1} = \Theta_{j,l-1} \cup SM_j$ is the set of possible models of agent j with a strategic level $l-1$ where $\Theta_{j,l-1}$ is the set of computable *intentional models* of agent j; SM_j is the complementary set of subintentional models of j. Intentional models of agent j : $\theta_{j,l-1} = \langle b_{j,l-1}, \hat{\theta}_j \rangle$, where $b_{j,l-1}$ is j's level $l-1$ belief; the *frame*, $\hat{\theta}_j = \langle A, \Omega_j, T_j, O_j, R_j, OC_j \rangle$.
- $A = A_i \times A_j$ is the set of joint actions of both agents.
- T_i is a transition function, $T_i : S \times A \times S \rightarrow [0, 1]$, which describes the results of the agents' actions on the physical states of the world.
- Ω_i is the set of agent i's observations.
- O_i is an observation function, $O_i : S \times A \times \Omega_i \rightarrow [0, 1]$, which gives the likelihood of perceiving observations in the state resulting from performing the action.
- R_i is defined as, $R_i : IS_i \times A \rightarrow \mathbf{R}$. An agent is allowed to have preferences over physical states and models of other agents, usually only the physical state will matter.
- OC_i gives the criterion for measuring optimality, which is typically the discounted sum of rewards in an extended interaction.

Agent i's level l belief $b_{i,l}$ is a probability distribution over i's interactive states $IS_{i,l}$ which contains not only physical states but agent j's possible models. Agent j's model $\theta_{j,l-1}$ contains agent j's level $l-1$ belief, $b_{j,l-1}$, a probability distribution over j's interactive states $IS_{j,l-1}$ which in turn is defined as $S \times M_{i,l-2}$, physical states and agent i's level $l-2$ possible models. This recursive defintion for agent beliefs continues until level 0 is reached where belief b_0 is a probability distribution over only physical states.

Previously, Doshi et al. (2010) modeled the individual three-stage Centipede games using I-POMDP$_{i,2}$. We extend this modeling to the four-stage games considered here using I-POMDP$_{i,3}$. As Doshi et al. noted, the decisions at the different points in the game are naturally modeled recursively rather than sequentially. The physical state space, $S = \{A, B, C, D, E\}$, is perfectly observable; i's actions, $A_i = \{Stay, Move\}$ are deterministic and j has similar actions; i observes other's actions, $\Omega_i = \{Stay, Move\}$; O_i is not needed; and R_i captures the diagnostic preferential ordering of the states contingent on which of the three games in a trial is being considered.

Given the opponent types used in the experimentation, intuitively, model set, Θ_j, contains three models of agent j (participant's opponent), that is $\{\theta_{j,0}, \theta_{j,1}, \theta_{j,2}\}$, where $\theta_{j,0}$ is the level 0 (myopic) model of the opponent, $\theta_{j,1}$ is the level 1 (predictive) model and $\theta_{j,2}$ is the level 2 (super-predictive) model. Predictions about the opponent's action by the participants were consistent with these three models being attributed. Parameters of these models of agent j are analogous to the I-POMDP for agent i, except for R_j which reflects the preferential ordering of the states for the opponent. The super-predictive model of agent j, $\theta_{j,2}$, includes the level 1 model of agent i, $\theta_{i,1}$, which includes the level 0 model of agent j, $\theta_{j,0}$, in its interactive state space.

Agent i's level 3 initial belief, $b_{i,3}$, assigns a probability distribution to its interactive state $IS_{i,3}$, which includes agent j's models based on the game being considered. This belief will reflect the general de facto thinking of the participants about their opponent. It also assigns a marginal probability 1 to state A indicating that agent i decides at that state. Agent j's beliefs $b_{j,0}$, $b_{j,1}$ and $b_{j,2}$ that are part of agent j's three models, respectively, assign a marginal probability 1 to B indicating that agent j acts at that state. Agent i's belief $b_{i,1}$, that is part of $\theta_{i,1}$, assigns a probability of 1 to state C. Additionally, agent j's belief $b_{j,0}$, that is part of $\theta_{j,0}$, assigns a probability of 1 to state D.

Previous investigations of strategic behavior in games, including Rosenthal's Centipede games, attribute social models such as reciprocity and altruistic behavior to others (McKelvey and Palfrey 1992; Gal and Pfeffer 2007). Reciprocity motivates participants to reward kind actions and punish unkind ones and is usually observed when the sum of payoffs for both players has a potential to increase through these actions. However, the setting of fixed-sum with no increase in total payoffs makes the games we consider competitive and precludes these models. Another social process that could explain some of the participants' behavior is a desire for inequality aversion (Fehr and Schmidt 1999), which would motivate participants to choose an action that leads to similar chances of winning for both players. For example, such a desire should cause participants to move proportionately more if the chance of winning at state A is 0.6 and this chance is preferentially in the middle, than say, when the chance at A is between 0.45 and 0.55. However, participants displayed a lower move rate of about 12% in the former case compared to a move rate of 14.5% in the latter

case. Hence, we believe that inequality aversion did not motivate a significant number of the participants.

Judgment and Decision Models From Fig. 2(a, b) and our analysis, notice that some of the participants learn about the opponent model as they continue to play. However, the rate of learning varies across participants, and, in general, the learning is slow and partial. This is indicative of the cognitive phenomenon that the participants could be underweighting the evidence that they observe. We may model this by making the observations slightly noisy and augmenting normative Bayesian learning in the following way:

$$Pr(\theta_{j,l}|o_i;\gamma) = \alpha Pr(\theta_{j,l})\ Pr(o_i|\theta_{j,l})^{\gamma} \quad (1)$$

where α is the normalization factor, l is the nested level of the model, if $\gamma < 1$, then the evidence $o_i \in \Omega_i$ is underweighted while updating the belief over j's models.

In Fig. 2(c), we observed significant rationality errors in the participants' decision making. Such noisy play was also observed by McKelvey and Palfrey (McKelvey and Palfrey 1992), and included in the model for their data. We utilize the *quantal response* model to simulate human non-normative choice. This model is based on the finding that rather than always choosing the optimal decision which maximizes the expected utility, individuals are known to select actions proportionally to their utilities. The quantal response model assigns a probability of choosing an action as a sigmoidal function of how close to optimal is the action. In the experiment, a rationality error occurs when a participant's action is not the best response to her prediction of the opponent's action.

Previously, Doshi et al. (2010) augmented I-POMDPs with both these models in order to simulate human recursive reasoning up to level two. As they continue to apply to our data, we extend the I-POMDP model to longer Centipede games and label it as I-POMDP$^{\gamma,\lambda}_{i,3}$.

The methodology for the experiments reveals that the participants are deceived into thinking that the opponent is human. Therefore, participants may justify unexpected actions of the opponent as errors in their decision making rather than due to their level of reasoning. Hence, we generalize the previous model by attributing quantal response choice to opponent's action selection as well. Let λ_1 be the quantal response parameter for the participant and λ_2 be the parameter for the opponent's action. Then,

$$Q(a_i^*;\gamma,\lambda_1,\lambda_2) = \frac{e^{\lambda_1 \cdot U(b_{i,3},\lambda_2,a_i^*)}}{\sum_{a_i \in A_i} e^{\lambda_1 \cdot U(b_{i,3},\lambda_2,a_i)}} \quad (2)$$

where parameters, λ_1, $\lambda_2 \in [-\infty, \infty]$; a_i^* is an action of the participant and $Q(a_i^*)$ is the probability assigned to the action by the model. $U(b_{i,3}, \lambda_2, a_i)$ is the utility for i on performing the action, a_i, given its belief, $b_{i,3}$, with λ_2 parameterizing j's action probabilities in the computation of the utility function. We label this model as I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$.

Learning Parameters from Data

In I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$, three parameters are involved: γ representing participants' learning rate, λ_1 and λ_2 representing non-normative actions of the participant and her opponent, respectively. The empirically informed I-POMDP model gives a likelihood of the experiment data given specific values of the three parameters.

We begin by learning λ_2 first. Because this parameter characterizes expected opponent behavior, we utilize the participants' predictions of their opponent's action in each game as the data. Denoting this set of predictions as $\mathcal{P}$, the likelihood of $\mathcal{P}$ is obtained by taking the product of $Q(a_{ij}^*;\lambda_2)$ over G games and N participants because the probability is hypothesized to be conditionally independent between games given the model and is independent between participants.

$$L(\mathcal{P};\lambda_2) = \prod_{i=1}^{N}\prod_{g=1}^{T} Q(a_{ij}^*;\lambda_2)$$

Here, a_{ij}^* is the observed prediction by participant i of opponent j's action, and $Q(a_{ij}^*;\lambda_2)$ is the probability assigned by the model to the action, whose computation is analogous to Eq. 2 except that j's lower-level belief replaces i's belief and j does not ascribe non-normative choice to its opponent.

In order to learn parameters, γ and λ_1 (or γ and λ in I-POMDP$^{\gamma,\lambda}_{i,3}$), we utilize the participants' actions at state A. Data consisting of these actions is denoted as $\mathcal{D}$. The likelihood of this data is given by the probability of the observed actions of participants i as assigned by our model over all games and participants.

$$\begin{aligned} L(\mathcal{D};\gamma,\lambda_1,\lambda_2) &= \prod_{i=1}^{N}\prod_{g=1}^{G} Q(a_i^*;\gamma,\lambda_1,\lambda_2) \\ &= \prod_{i=1}^{N}\prod_{g=1}^{G} \frac{e^{\lambda_1 \cdot U(b_{i,3}^g,\lambda_2,a_i^*)}}{\sum_{a_i \in A_i} e^{\lambda_1 \cdot U(b_{i,3}^g,\lambda_2,a_i)}} \quad \text{(from Eq. 2)} \end{aligned}$$

We may simplify the computation of the likelihoods by taking its log:

$$\begin{aligned} LL(\mathcal{P};\lambda_2) &= \sum_{i=1}^{N}\sum_{g=1}^{T} log\ Q(a_{ij}^*;\lambda_2) \\ LL(\mathcal{D};\gamma,\lambda_1,\lambda_2) &= \sum_{i=1}^{N}\sum_{g=1}^{G} log\ Q(a_i^*;\gamma,\lambda_1,\lambda_2) \end{aligned} \quad (3)$$

To estimate the values of the three parameters (γ, λ_1, λ_2) in our empirically informed I-POMDP model, we maximize the log likelihoods in Eq. 3 using the Nelder-Mead simplex method (Nelder and Mead 1965). Notice that the ideal Q functions will assign a probability of 1 to the observed actions resulting in a log likelihood of zero, otherwise the likelihoods are negative.

Model Performance

We use stratified, five-fold cross-validation to learn the parameters and evaluate the models.

Parameters In order to learn λ_2, we use the prediction data for the catch games only. This is because no matter the type of the opponent, the rational action for the opponent in catch games is to move. Hence, predictions of stay by the participants in the catch trials would signal a non-normative action selection for the opponent. This also permits learning a single λ_2 value for participants in the three groups. However, this is not the case for the other parameters. In Fig. 2, we observe that for different opponents, the learning rate, L, is different. Also, in Fig. 2(c), we observe that the rationality errors differ considerably between the opponent groups. Therefore, we learn parameters, γ and λ_1 given the value of λ_2, seperately from each group's diagnostic games. We report the learned parameters averaged over the five folds in Table 1.

param.	myopic	predictive	super-predictive
λ_2	1.959		
γ	0.297	0.081	0.305
λ_1	3.254	3.892	3.785

Table 1: Parameter values learned from the experiment data for I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$.

From Table 1, we see that γ for the predictive opponent group is close to zero. This is consistent with the observation that participants in this group did not make much progress in learning the opponent type. Consequently, we focus our analysis on the myopic and super-predictive opponent groups here onwards.

model	log likelihood	
	myopic	super-predictive
Random	-1455.605	-1414.017
I-POMDP$^{\gamma,\lambda}_{i,3}$	-732.401	-437.6919
I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$	-706.377	-431.4761

Table 2: Log likelihood of the different models with parameters as in Table 1.

We show the average log likelihoods of the different models, including a random one that predicts other's actions randomly and chooses its own actions randomly, in Table 2. The random model serves as our null hypothesis. We point out that I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$ has the highest likelihood among all three models, although the likelihoods of the other I-POMDP based model is not much different especially for the super-predictive group.

Achievement and Prediction Scores We utilize the learned values in Table 1 to parameterize the underweighting and quantal responses within the I-POMDP based models. We cross-validated the models on the test folds. Using a participant's actions in the first 5 games, we initialized the prior belief distribution over the opponent types. The average simulation performance of the I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$ model is displayed in Fig. 4.

As we see in Fig. 4, model-based achievement and prediction scores have similar values and trends as the

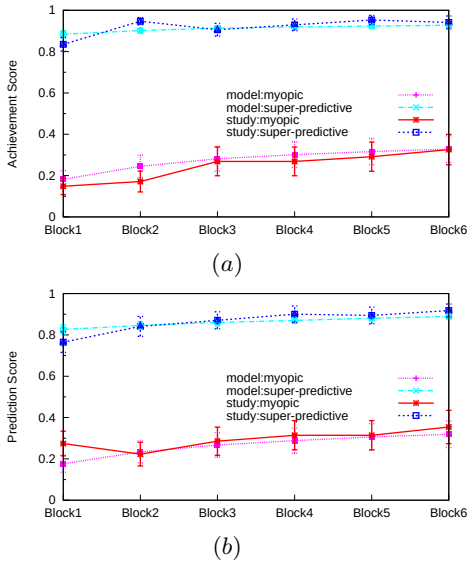


Figure 4: Comparison of model predictions with actual data in test folds: (a) Mean achievement scores and (b) Mean prediction scores.

experiment data. However, there is some discrepancy in the first block. This is caused by the difficulty in determining an accurate measure of the participant's initial beliefs as they started the experiment. Each model data point is the average of 500 simulation runs.

We also measure the goodness of the fit by computing the mean squared error (MSE) of the output by the models, I-POMDP$^{\gamma,\lambda}_{i,3}$, I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$, and compare it to those of the random model (null hypothesis) for significance. We show the MSE for the achievement and prediction scores based on the models in Table 3.

Notice from Table 3, that both I-POMDP based models have MSE that are significantly lower than the random model. Although the log likelihood of I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$ is higher than the other for both opponent groups (see Table 2), MSE values do not significantly distinguish one model over the other across both groups. Hence, the improvement in log likelihood (small as it is) does not result in a better simulation performance for I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$ across both scores and opponent groups. While attributing non-normative action selection to the opponent results in more accurate predictions for the super-predictive opponent, it does not translate to actions that better models the observed data. This could be because of some inconsistency between the participants' predictions of the opponent actions and their own actions possibly due to inattention to the prediction screen. As such, our results do not conclusively distinguish between the performance of the two models although both fit well.

Opponent type	Mean Squared Error (MSE)					
	Achievement score			Prediction score		
	Random	I-POMDP$^{\gamma,\lambda}_{i,3}$	I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$	Random	I-POMDP$^{\gamma,\lambda}_{i,3}$	I-POMDP$^{\gamma,\lambda_1,\lambda_2}_{i,3}$
myopic	0.01170	0.00248	0.00139	0.00316	0.00157	0.00200
super-predictive	0.34458	0.00076	0.00094	0.28529	0.00142	0.00098

Table 3: Goodness of the fit of different models with the study data.

Discussion

An alternate explanation of the high achievement score in super-predictive group could be that participants employed backward induction (or minimax) to solve the game instead of recursively thinking about opponent behavior. However, significant achievement scores in the myopic group indicates that the participant pool likely did not apply these methods. Also, presence of learning of opponent models as subjects played the games provides further evidence of recursive reasoning. Finally, Goodie et al. (2010) report that a debriefing questionnaire in prior experiments did not reveal evidence that the population was aware of these techniques.

The models that we apply here fall within the class of belief learning models as classified by Camerer (2003). Other belief learning models have enjoyed good empirical support as well, one of which is the weighted fictitious play. This model maintains a distribution over possible actions of the opponent and updates it based on the frequency of the observed actions. Hence, it represents a simple way to update beliefs and could apply here. We plan to compare our existing models with this one as part of future work.

Acknowledgement

Research reported in this paper was supported in part by a grant FA9550-08-1-0429 from AFOSR and by a NSF CAREER grant IIS-0845036. We would also like to acknowledge the helpful comments from anonymous reviewers.

References

Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Doshi, P.; Qu, X.; Goodie, A.; and Young, D. 2010. Modeling recursive reasoning in humans using empirically informed interactive pomdps. In *International Autonomous Agents and Multiagent Systems Conference (AAMAS)*, 1223–1230.

Fehr, E., and Schmidt, K. 1999. A theory of fairness, competition and cooperation. *Quaterly Journal of Economics* 114:817–868.

Ficici, S., and Pfeffer, A. 2008. Modeling how humans reason about others with partial information. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 315–322.

Gal, Y., and Pfeffer, A. 2007. Modeling reciprocal behavior in human bilateral negotiation. In *Twenty-Second Conference on Artificial Intelligence (AAAI)* 815-820.

Gmytrasiewicz, P., and Doshi, P. 2005. A framework for sequential planning in multiagent settings. *Journal of Artificial Intelligence Research* 24:49–79.

Goodie, A. S., Doshi, P. and Young, D. L. 2010. Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making* DOI:10.1002/bdm.717

Hedden, T., and Zhang, J. 2002. What do you think I think you think?: Strategic reasoning in matrix games. *Cognition* 85:1–36.

McKelvey, R. D., and Palfrey, T. R. 1992. An experimental study of the centipede game. *Econometrica* 60:803–836.

McKelvey, R., and Palfrey, T. 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior* 10:6–38.

Nelder, J.A., and Mead, R. 1965. A simplex method for function minimization. *Computer Journal* 7:308–313.

Rosenthal, R. 1981. Games of perfect information, predatory pricing and the chain store paradox. *Journal of Economic Theory* 25:92–100.

Stahl, D., and Wilson, P. 1995. On player's models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10:218–254.