Humans depart from optimal computational models of socially interactive decision-making under partial information

Saurabh Steixner-Kumar[1,⋆], Tessa Rusch[1,2], Prashant Doshi[3], Jan Gläscher[1,⋆,†], Michael Spezio[4,1,⋆,†]

1 Institute of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Germany
2 Division of the Humanities and Social Sciences, California Institute of Technology, CA, USA
3 Department of Computer Science, University of Georgia, GA, USA
4 Psychology, Neuroscience, and Data Science, Scripps College, CA, USA

† These authors contributed equally to this work.

* Correspondence: s.steixner-kumar@uke.de, glaescher@uke.de, mspezio@scrippscollege.edu

Abstract

Decision making under uncertainty and under incomplete evidence in multiagent settings is of increasing interest in decision science, assistive robotics, and machine assisted cognition. The degree to which human agents depart from computationally optimal solutions in socially interactive settings is generally unknown. Yet, this knowledge is critical for advances in these areas. Such understanding also provides insight into how competition and cooperation affect human interaction and the underlying contributions of Theory of Mind. In this paper, we adapt the well-known 'Tiger Problem' from artificial-agent research to human participants in single agent and interactive, dyadic settings under both competition and cooperation. A novel element of the adaptation required participants to predict the actions of their dyadic partners in the interactive Tiger Tasks, to facilitate explicit Theory of Mind processing. Compared to computationally optimal solutions, participants gathered less information before outcome-related decision when competing with others and collected more evidence when cooperating with others. These departures from optimality were not haphazard but showed evidence of improved performance through learning across sessions. Costly errors resulted under conditions of competition, yielding both lower rates of rewarding actions and lower accuracy in predicting the actions of others, compared to prediction accuracy in cooperation. Taken together, the experiments and collected data provide a novel approach and insights into studying human social interaction and human-machine interaction when shared information is partial.

1 Introduction

Knowing what contributes to successful cooperation and competition is critical for ensuring organizational and institutional sustainability in flourishing societies. This knowledge is required by teams of investors, businesses, manufacturing design and implementation, health care workers, jurors, security forces, search-and-rescue specialists, government agencies, and even voters. Each person or group seeks evidence gained from limited, partial observations, to reach a decision: Is it time to act on Option A or Option B, or is better to keep accumulating evidence before performing a consequential and irrevocable action?

In each context, formal computational models of agent actions can identify optimal sequences of exploration and/or option selection. Ideally, these models will allow for robust artificial intelligence (AI) systems that pair with human agents in contexts of competition and cooperation. These contexts include socially assistive robotics [1, 2, 3] and machine assisted cognition or decision making [4, 5, 6, 7]. To achieve these goals, principled design for machine-assisted cognition must move beyond modeling normative actions and action probabilities, as noted by Pacaux-Lemoine and Flemisch [8], to develop accurate models of the beliefs and intentions of the human agents. Only following such a research program will yield the best adaptations of machine-based guidance in shifting contexts. Models of beliefs and intentions require accounting for how human agents' beliefs, intentions, and actions differ from computationally optimal solutions, since what is computationally optimal is only rational under the conditions of narrow, static value functions and a limited array of learning algorithms. One especially needs to know how these differences are affected by changes in the competitive and cooperative environments which in turn influence human state representation and valuation of gains and losses for self [9, 10] and others [11]. Progress in meeting these requirements will involve studying how human agents act in the same simulated partially observable, valuationally uncertain tasks that are used to advance robotic and other computational agents in multiagent interaction.

This paper adapts the "Tiger Problem", a widely known artificial agent simulation framework for single agent and multiagent modeling under conditions of uncertainty and partial observation [12, 13], to demonstrate influences of competition and cooperation on choice and choice timing of human participants in interaction. We have adapted the Tiger Problem to create several Tiger Tasks in single and multiagent, interactive formats. We hypothesized that human agents will depart significantly from computationally optimal actions. While we focus here on presenting outcomes and optimal models of the actions of single agents and of dyadic interaction in our single agent Tiger Task (TT) and Interactive Tiger Tasks (ITTs) respectively, our future work will pursue formal modeling of the beliefs and intentions of human agents and of the agents' beliefs about the beliefs and intentions of other agents.

Decision-making under uncertainty has been a central focus of decision science since its inception [14]. Several computational models have been developed to provide normative solutions to this problem and to accurately model empirical data in such situations. Most of these postulate an *expected reward* as a common currency [9, 15, 16] on which all decision options are projected and upon which choices are made. At its core, such an expected value combines information about the probability and magnitude of an associated outcome. By gathering more evidence, the uncertainty about these two properties can be reduced and decisions can improve.

The fields of operations research and of artificial intelligence in computer science have long investigated challenges of this kind [17, 18, 19]. In a seminal paper, Kaelbling et al [12] introduced a novel simulation problem environment and model for investigating decisions in which an agent needs to develop beneficial strategies or policies when they only have partial information about the state of the environment. The "Tiger Problem" places an agent in a situation in which there is a pot of gold incurring a small monetary gain and a tiger incurring a large monetary loss, each of which is behind one of two doors. The agent's task is to listen at the doors for the tiger's growl, which is partial evidence of the tiger's location. When the agent is sure about the location of the tiger, they should open the other door to receive the reward. Repeated trials allow the agent or agents to build up beliefs regarding the tiger's (and the reward's) location. On each trial the agent chooses to gather

more information by choosing Listen, or chooses to take a consequential action by choosing either Open Left Door or Open Right Door. Weighing the gathered evidence against the potential gain/loss is crucial for decision making under partial information.

One of the limitations of the original Tiger Problem is that it was incapable of addressing multiagent contexts in which agents must develop models and be sensitive to other agents' representations of the environment. These contexts require some way of formally modeling the models of other agents, that is, modeling a "theory of mind" (ToM; [20]), especially in situations where both the environment and other agents' actions constitute critical uncertainties in decision making. To overcome this limitation Gmytrasiewicz and Doshi [13] introduced the interactive Tiger Problem (ITP), along with a modeling solution. In the interactive case, each agent can make probabilistic observations regarding both the tiger and the other agents' actions. When agents in the ITP listen, they listen for both growls and creaky doors vs. silence. If a growl comes from the door on the left, that is partial evidence that a tiger is behind that door. If a creak comes from the door on the right, that is partial evidence that the other agent opened that door. If no door makes a creaking sound, that is evidence that the other agent is also listening.

The conditions of competition and cooperation are expected to lead human agents to depart from computationally optimal actions. In competitive contexts, agents may experience pressure to race one another to the door that has the reward. Competitors may act more hastily, prior to obtaining all of the evidence they need, since they will weigh reward against the risk of opening the wrong door and the risk of one's competitor opening the correct door. In cooperative contexts, agents may decide to take more time to listen since they expect others will also patiently gather as much evidence as possible to ensure a maximum reward for all.

The goals of this study were to adapt the Tiger Problem from AI to the domains of human action and interaction, under both cooperative and competitive contexts. We adapted the Tiger Problem and the Interactive Tiger Problem to a single-agent Tiger Task (TT) and a dyadic Interactive Tiger Task (ITT), respectively. Using modifications to the reward or payout matrices, following [21], we presented participants either with a competitive or a cooperative ITT. This allowed us to compare human choices in the individual Tiger Task (TT) and in both the competitive and cooperative variations of the Interactive Tiger Task (ITT) to what we already know to be computationally optimal choices. In our ITT design, we also added a novel aspect, requiring participants to predict the action of the other agent prior to choosing their own action. Immediately prior to reporting their own choice in the competitive or cooperative version of the ITT, each participant needed to report their prediction of the other participant's choice. We introduced this element as a way of facilitating explicit focus on the other person and their actions, which increases the likelihood of participants engaging ToM processing [22, 23]. This step should elicit data that allows us to model and assess putative ToM processes more directly in our future work.

In this study, we wanted to identify how cooperation and competition influence departures from theoretical optima, and to compare these departures directly to the departures from optimality in the single-agent TT. This approach allows us to address several of the key requirements for optimizing socially assistive robotics and machine assisted cognition, as discussed. Further, we wanted to assess whether any departures from computationally optimal choices might benefit human agents by associating with better performance in avoiding the tiger and reaching the reward, and in more accurate predictions of other agents' actions. Machine assisted cognition and socially assistive robotics may fail

in their goals if they do not adequately model what real human agents actually value, believe, and do. Computational optimality will in fact become suboptimal with respect to the goals of assistive robotics if the recommended interactive options result only from static, artificial value functions and do not adapt to human agents' departures from those functions. Situations could also arise in which human agents' expertise is important. In complex multiagent interactions for which human agents have some expertise, human situated cognition may yield values, beliefs, and actions that end in more optimal outcomes than those driven by limited value functions and learning algorithms, especially under conditions of limited data availability. Even in a simplified task context such as the ITT, we would expect that human agents can produce accurate predictions of other agents' actions, if human agents did in fact model other agents' beliefs and expected values. Thus, we present analyses of overt predictions of others' choices as an approximate readout of ToM processes, and we assess how the accuracy of those predictions depends upon choices during cooperation and competition between persons.

We test and report two versions of the reward matrices and our hypotheses are as follows. We do not expect the differences between the two reward matrices to strongly affect participant actions, while of course expecting differences in computationally optimal behavior. Our original reward matrix [12, 13] presents a large risk/reward ratio. Under this large ratio of risk to reward, we expect participants in the TT to listen for the tiger much less than is optimal because we expect that humans would overweight losses from sequential listening actions (i.e., -1) and underweight the probability of loss when opening the wrong door. When the risk/reward ratio is much smaller, as in our second, modified reward matrix, we expect participants in the TT to listen for the tiger optimally or slightly more than is optimal. The modified reward matrix in the TT is more forgiving of errors than is the original matrix and keeps the penalty for listening the same. In the ITTs, we expect that competition and cooperation would lead to divergent response patterns. Under competition, we expect that people would race the other participant to the good door, and so would listen less than they do in the TT, despite the ITTs requiring additional cognitive load to integrate partial observations of others' actions (i.e., creaks, silence) than is the case in the TT. Conversely, we expect cooperation to elicit greater care and coordination in evidence gathering so that both participants would have the best opportunities to maximize their joint rewards. So participants in the cooperative ITT were expected to listen more often than in the competitive ITT and in the TT.

Since participants in the cooperative ITT are expected to listen more and so to have greater access to evidence, and since better evidence should include estimation of others' models and actions, we expect cooperation to yield more accurate predictions about other agents' actions, compared to competition. We also expect that people would choose actions consistent with their predictions more often during cooperation than during competition, suggesting greater explicit or implicit confidence in the evidence underlying those predictions or in the predictions themselves.

2 Methods and Materials

2.1 Participants

This study was approved by the Ethics Committee of the German Psychological Association (ref no. JG012015-052016) and carried out according to the Declaration of Helsinki. All participants gave written informed consent and were financially reimbursed for their participation. We invited 182 participants to play several variants of the Tiger Task (TT and ITT), all of whom were naive to the

task. The variants differed in their payout structure (original/modified, see below for details), their complexity (single/multiagent), and their interactive context (cooperative/competitive). From the total, 58 participants played the original payout structure variant of the TT [12], after which they played in either the cooperative (30 participants) or the competitive context (28 participants) with random assignment. All participants played the single-agent version of the TT before the multiagent ITT version. The remaining 124 participants took part similarly in the modified payout structure variant of the TT and ITT. Here after participating in the TT, half the participants played (randomly assigned) in the cooperative and the other half in the competitive context.

2.2 Experimental Schedule

An ITT dyad consisted of 2 participants who were comfortably seated next to each other at two computer screens. A partition separated the participants so that each could only see their own screen. The experimenter was a muted observer for the duration of the task, separated from the two participants by an additional partition (Figure 1). The room temperature averaged $mean \pm s.d.$: 22.59 ± 1.31 degrees Celsius. The dyad interacted using either the original payout structure or the modified payout structure of the ITT after participating in the respective TT. Each dyad played either in the cooperative or the competitive context, but not both. The experimental setup was implemented in the Psychtoolbox Version 3.0.14 running under MATLAB version 9.1.0 (The MathWorks, Natick, MA). Concurrent dyadic, synchronized high density EEG was recorded from a subset of participants, for use in subsequent analyses. Those will not be presented in this paper. Good performance also earned participants a relative financial bonus. The experimental code for running the experiment is available from the project's github page `https://github.com/SteixnerKumar/tiger_task_experiment`.

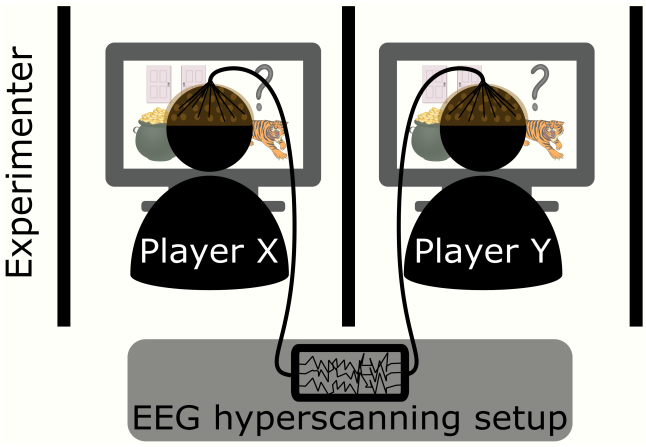

Figure 1: Task schematic of the experiment. The dyad (two participants X and Y) were separated by a partition. Each participant faced their own computer screen playing the tiger task and wore headphones playing constant static to ensure that they could not hear key presses or other low level sounds. Additionally, each wore an EEG cap in a hyperscanning setup. The experimenter silently monitored the behavior and EEG from behind another partition.

2.3 Task

In the task, individual trials began with two doors presented on the computer screen. Participants understood that each door concealed either a tiger or a pot of gold (i.e., reward). The task was to open the correct door (i.e., the one hiding the pot of gold) and avoid the tiger. On each round, a

participant had a set of three different actions available: listen (L), open the door on the left (OL), and open the door on the right (OR). The L action gave a probabilistic hint about the location of the tiger that is 70% accurate, via a growl behind the left door (GL) or a growl behind the right door (GR). The OL/OR actions opened the chosen door. After the participant opened a door, they saw the result and the system randomly reallocated the location of both tiger and reward (50% chance of being behind any particular door). Participants understood the underlying probabilities via task instructions and saw the potential reward associated with each action during each round, via the payout matrix. (refer to the figures 2 and 3 for a visual guide). The task also differed in complexity depending on whether it was the single-agent version (TT) or the multiagent version (i.e., ITT). A "session" of the TT or ITT was a total of 10 tiger-trials (i.e., 10 door openings). The participant sought to maximize the total rewards obtained during the particular task session. In the ITT, L actions provided probabilistic information about both the tiger's location and the other agent's action. The sound of a creaking door on the left or right (CL or CR) gave partial evidence that the other agent opened the corresponding door and silence (S) gave partial evidence that no door was opened. These cues were 80% accurate regarding the true action of the other agent. Additionally, the tasks used color coding to help differentiate actions of self (yellow) and other (blue).

2.3.1 Single-agent Version

In the single-agent version, the participant played alone. A "tiger-trial" consisted of a sequence of listen actions that ended with the participant opened one of the doors. During a session, the location of the tiger/reward changed only if a door was opened.

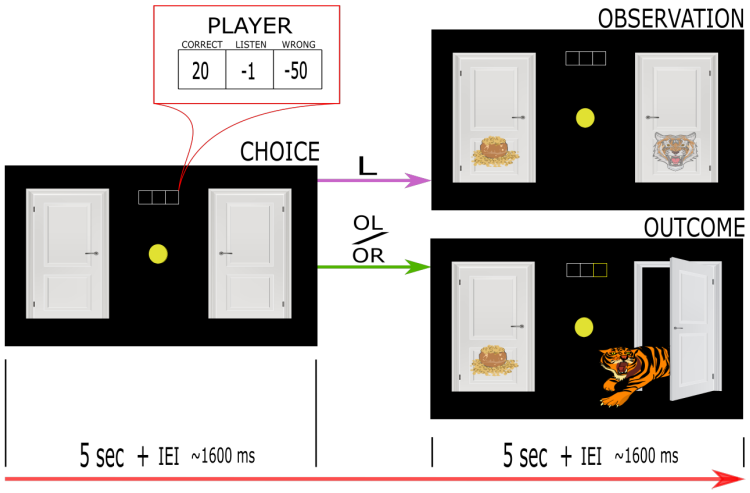

Figure 2: Task sequence of the single-agent version of the tiger task (TT). At the beginning of each decision step, the participant sees two doors and is tasked to make an action choice (CHOICE) from a set of three actions (Listen (L), Open-left (OL), Open-right (OR)). Participants have 5 seconds to make their choice, else a default listen action is chosen for them. A chosen listen action leads to a probabilistic observation (PHYSICAL-OBSERVATION, 70% correct), and the sequence repeats. The probabilistic hints are signified by semi-transparent images of the tiger and the gold-pot on the respective doors. There is always an inter-event-interval (IEI) of $\sim$ 1600 milliseconds. Choosing left/right open action opens the door (OUTCOME), where the participants either receive the gold-pot or encounter the tiger (tiger encounter is shown here). Opening the door resets the tiger and the gold-pot to begin the task sequence again. The matrix in the red box shows the potential payouts the participant can expect upon the different action choices (constantly displayed through the TT).

The task began with a choice-screen (5*sec.*) asking the participant to choose an action (Figure 2). The number of L actions were not limited, but if a choice was not made during the allotted time, a forced listen action was assumed to move the task forward. Following an intra-event-interval (IEI)(+ $\sim$ 1600*ms.*), either the observation-screen was displayed (5*sec.*) after a listen action (L), or the outcome screen (5*sec.*) was shown after an open-left (OL)/open-right (OR) action, which revealed the gain from the particular tiger-trial and the total gain from the session to that point. Opening the correct door (pot of gold) incurred a small win, whereas opening the incorrect door (tiger) incurred a large loss. The participant gathered sufficient evidence about the tiger location through a series of L actions (incurring a small loss each time) before opening a door (OL/OR action).

2.3.2 Multiagent Version

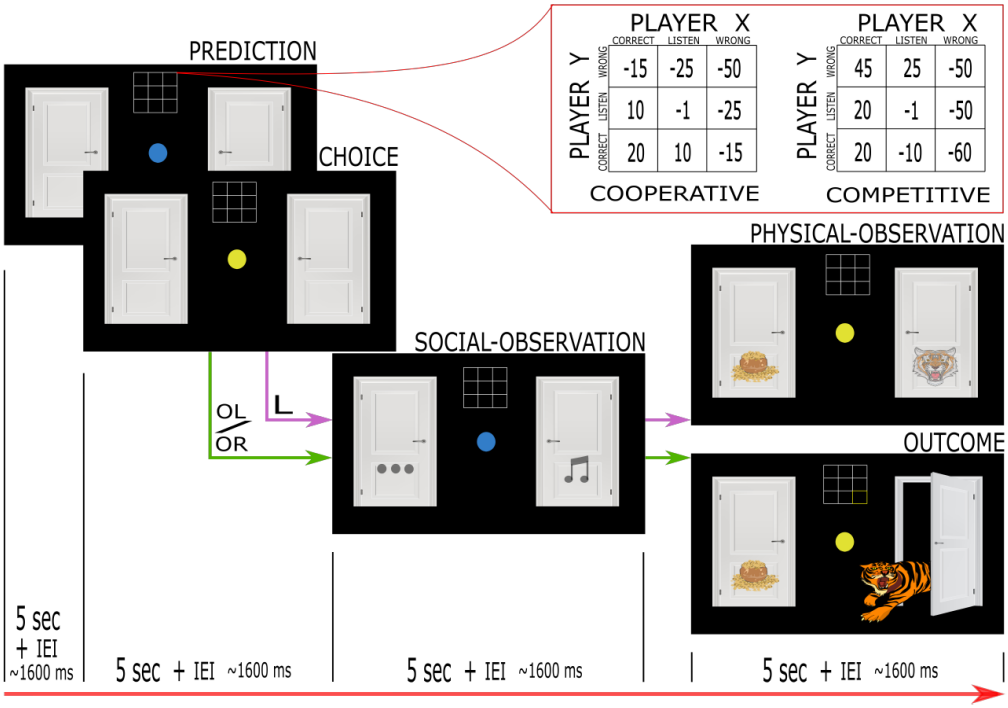

Figure 3: Task sequence of the multiagent version (cooperative and competitive context) of the interactive tiger task (ITT). The task sequence is color-coded for the participant. Blue represents the other participant and yellow represents the self. To facilitate attention and Theory of Mind regarding the other agent, the participant begins each round of the task by predicting the action choice of the other participant (PREDICTION, the blue dot in the middle indicating that the agent should think about the other agent's action. Possible actions: L, OL, OR). The participant has 5 seconds, else the default L is chosen to move the task forward. After an inter-event-interval (IEI) of $\sim$ 1600 milliseconds, the participant has to make the personal action choice (CHOICE, represented by the yellow dot in the middle indicating that the agent should think about their own action). For this the participant similarly has 5 seconds, else the default action L is chosen. Regardless of the one's own action, the participant obtains probabilistic (80% correct) social information (SOCIAL-OBSERVATION, the blue dot in middle) about the other participant before either receiving the probabilistic hint (70% correct) about the locations of the tiger and the pot of gold. (PHYSICAL-OBSERVATION, yellow dot in the middle of the screen) for L actions, or opening the door (OUTCOME, yellow dot in the middle of the screen) for OL or OR actions, to get the gold-pot/tiger (tiger encounter is shown in the figure). Opening the door randomly reallocates the tiger and the reward locations to begin the tiger-trial sequence again. The matrices in the red box show the potential payouts the participant can expect as determined by the choices made by the agent (participant X) and the other agent (participant Y). This is context-specific for cooperation and competition (constantly displayed through the ITT).

The ITT consisted of both cooperative and competitive contexts. To perform optimally in the task, participants had to use partial observations to update beliefs about the tiger's location (i.e., growls (GL/GR)) and the other participants' actions (i.e., creaks (CL/CR), silence (S)), along with representations of those participants' beliefs and expected values. This medley made the ITT complex because the rewards were determined by joint actions of both participant, which required higher cognitive loads and social-cognitive processes to develop a strategy for when best to open a door. In the cooperative context, the maximum joint reward occurred when both participants opened the correct door on the same round, while in the competitive context, the maximum payoff to a participant was obtained when the participant opened the correct door while the other participant opened the wrong door on the same round.

We modified the original design of the multiagent Tiger Problem [13] (Figure 3) such that each round began with the prediction-screen (with blue-dot) that required participants to predict the other participants' actions (L, OL or OR). Next, the choice-screen (with yellow-dot) prompted the participant for their own action (L, OL or OR), just as in the single-agent version. The participant then received a probabilistic hint about the other participants' actions (80% accurate) on the social-observation screen (with blue-dot). Finally, only after the partial observation of the other's action did the participant receive a payout in the outcome-screen (for OL/OR action choice) or a probabilistic hint (for L action choice) about the tiger's location, in the physical-observation screen (with yellow-dot). A "tiger trial" ended when either of the participants opened the door. Prediction, choice, social-observation, and physical-observation/outcome (all $5sec.$ duration) screens were all separated by the inter-event-interval ($1sec.+ \sim 600ms.$).

2.3.3 Differences in the Original and Modified TT

The original payout structure is taken from Doshi and Gmytrasiewicz formulation of the multiagent Tiger Problem and their solutions using POMDP and I-POMDP frameworks [21, 12]. In the single-agent version, an L action costs the participant -1 point (Figure 4). This is comparatively a small price to pay to get the reward of $+10$ points on getting the gold-pot. However, encountering the tiger punishes the participant quite harshly by taking away -100 points. The payout structure of the multiagent version of the task is envisioned such that the participants get their reward plus or minus half the reward of the other in the cooperative and competitive context respectively.

Verbal and written reports from participants completing the ITT with the original payout structure highlighted the demotivating effect of opening the wrong door. That is, the losses were so large that some participants thought they could never make up for one loss with just such small gains for opening the correct door on subsequent rounds. Therefore, we modified the payout structure in the following way: gains are doubled, and losses are halved. This can be easily recognized, except for the L action, which still incurs a cost of -1 point (Figure 4). Doubling the gain ($+20$ points), and halving the losses (-50 points) decreases the severity of the punishment and the loss-to-gain ratio, thus keeping the participants more engaged. The payouts of the ITT were also changed accordingly (see section 2.4 below), and we used whole numbers to make it easier for the participants to keep track of rewards and reward totals.

Besides the differences in the payouts, in the modified TT, the participants were trained in the single-agent version of the task for 10 trials, and on the ITT (cooperative or competitive, depending on random assignment) for 20 trials.

2.4 Payouts

We will describe the payout structure of the modified TT here. To make it simpler, the payout structure in Figure 4 is for the scenario where the tiger is behind the left door; Changing the rows and columns of OL and OR would represent the tiger behind the right door. For the sake of convenience, we will call the two participants X and Y. We will refer to a particular cell in the payout matrix by its row and column number; As an example, row 1 column 2 would be {1,2}.

Payout in the single-agent version is straightforward, where the listen action costs the participant a single point {1,1}. When the participant opens the correct door {1,3} the reward is +20 points while encountering the tiger {1,2} incurs a loss of -50 points.

Payouts in the modified matrix in the ITT departed from the original matrix, which are calculated by adding (in the cooperative context) or subtracting (in the competitive context) half of the single-player payouts (see Figure 4A). In contrast, modified payouts in the cooperative context of the multiagent version are completely symmetric to foster cooperation via shared outcomes. When the participants gather evidence through a listen action {1,1}, they lose −1 points each. The optimum scenario is when both open the correct door on the same round {3,3}, gaining the maximum +20 each. Other scenarios are relatively sub-optimal which incur losses. These moderate losses motivate participants to learn and look for the right door in subsequent trials. The worst scenario is when both participants get the tiger, losing −50 points each.

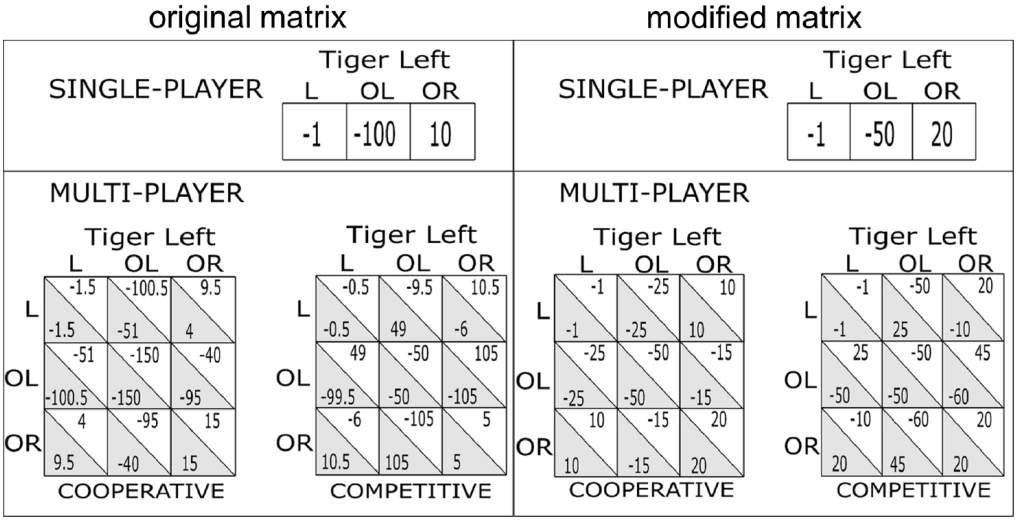

Figure 4: Original and modified payout structures of the TT and ITTs. The payouts are potential points that can be gained for a chosen action (listen (L) and open-left/right (OL/OR)), when the tiger is behind the left door. The points scheme remains the same for the tiger behind the right door if we switch the OL and OR columns and rows. In the original matrix for the single-participant setting, a L action costs -1 points. Getting the gold-pot rewards +10 points while encountering the tiger takes away -100 points. In the multiagent setting, the column actions represent one's own actions, while the row is the other participants' actions. The point system is similar from the single-participant setting but more combinations are added as the other participants' actions affect the points a participant can make. Depending upon the context, the points a participant gains are their own points plus half the points of the other participant (cooperative context) or minus the points of the other participant (competitive context). In the modified matrix single-participant setting the gains are doubled to +20 points and the losses are halved to -50 points. The multiagent setting has simpler whole numbers and the differences in the points is comparatively less extreme. To facilitate more cooperation in the cooperative context, the payout in the modified matrix is completely symmetric.

The main diagonal in the competitive context payout structure is similar to the cooperative one. Both participants' gathering evidence costs each −1 points. The best-case scenario for X is to open the correct door when Y opens the wrong door {2,3}; This results in a big win (+45 points) for X, along with a bad loss (−60 points) for Y. The payout structure also incentivizes a participant if the other is wrong, fostering competition to be the first one to the correct door.

2.5 Indices of behavioral performance

Our goal in this paper is to characterize participants' performance on the TT and on the ITT in two interactive contexts, cooperation and competition. We do this in terms of assessing the sequences and proportions of participants' choices, along with their predictions about the actions of the other participant. We also aim to understand how other-regarding predictions relate to an agent's choices by correlating the prediction and choice data. Finally, we also want to reveal learning-related changes and the ensuing improvement in task performance as both participants in the ITT learn about each other's choices and make better predictions.

We derived several behavioral indices from the choice and the prediction data of the ITT. In this section, we explain how they relate to the interactive contexts and what ToM processes they suggest.

Number of listen actions. One important characteristic of choices in both the TT, but more so in the ITT, is the degree to which human agents seek information prior to acting. We operationalize this via the number of listen actions prior to an open action, which concludes a tiger-trial. In this respect, the number of listen actions can be interpreted as a data-driven index weighing uncertainties of reward with those of failure/loss. The number of listen actions should follow the incentivizing structure of the payout matrices in the different interactive contexts. Using I-POMDP modeling with the multiagent Tiger Problem [21, 24], allows us to determine optimal numbers of L actions depending upon two key parameters of I-POMDP models: the ToM level and the planning horizon (see work by Doshi [24] for complete modeling details and definitions of level and planning horizon). In brief, ToM "level" (see [20]) applies to the level of recursion of Agent A's model of Agent B's beliefs, intentions, values, etc. Level 0 ToM is defined as no ToM at all. Level 1 ToM is defined as basic ToM with no recursion, such that Agent A does have a model of Agent B's beliefs, intentions, values, but has no model of Agent B's model of Agent A. Level 2 ToM is defined as a recursive model, in which Agent A has a model of Agent B's beliefs, intentions, values along with a model of Agent B's model of Agent A's beliefs, intentions, values, etc. Levels of ToM above 2 are higher levels of recursion.

The planning horizon in such models refers to number of iterations from the current round whose estimated outcome probabilities influence the choice of action. Horizon 1 means that the model takes account only of the current round's expectation. Horizon 2 means that the model looks one iteration ahead, and so on. We apply previously established models to provide optimal numbers of L actions for both the "Level 1, Horizon 1" (L1H1) and "Level 1, Horizon 2" (L1H2) models.

Evidence difference. Computational modeling reveals that several L actions are needed to gradually build up a belief about the tiger's location because the physical observations in the TT (tiger growls) are only partial observations of the tiger's true location and are only 70% accurate. Thus, the average evidence regarding the tiger's location prior to an agent's open action operationalizes that agent's evidence threshold and allows an estimation of deviation from computationally optimal evidence accumulation. We calculate the evidence difference as the number of observations from the true

location of the tiger minus the number of observations from the other side (e.g. if the tiger is on the left (TL), the evidence difference is calculated as (nGL|TL) − (nGR|TL), where n stands for number and | for conditionality). Compared to the number of Listen actions the evidence difference effectively controls for the inconsistency in an observation sequence that arises from the probabilistic nature of the physical observations. We calculate the evidence difference for every tiger-trial and average them to obtain the subject-specific evidence threshold before an open action is committed.

Identical open actions. This measure provides the most direct link to the incentivizing structure of the payout matrices for the interactive contexts. During cooperation, the best outcome is to open the correct door together and the worst outcome is to open the tiger's door together. During competition though, it is best to make it to the correct door first, and even better if the other person chooses the wrong door. So competition includes disincentives to identical open actions. Because they are such a direct expression of contextual differences in the ITT, identical open actions lend themselves particularly well for demonstrating learning-related improvements of joint performance on the ITT.

Correct open actions. Wins or losses are the results of correct and incorrect open actions in the TT and ITT. The number of correct open actions is therefore a measure of how well human agents have understood and carried out the task, along with the quality of evidence on which those agents base their actions.

While the above-mentioned measures characterize aspects of participant choice, their prediction data on the ITT provide indirect assessment of their ToM processes. A requirement for the most successful task performance in the ITT is that agents build mental models of the other participant, which they can query implicitly or explicitly to predict the other participant's next action. To partially access and assess these representations, we used several behavioral indices from the prediction data.

Number of listen predictions. Paralleling the number of listen actions, the number of *predicted* listen actions or the number of listen predictions indicates how much evidence a participant thinks the other participant needs before committing an open action. Just as with the number of listen actions above, the number of listen predictions can be interpreted as the expected risk sensitivity of the other participant. Similarly, if the number of listen actions is expected to be higher during cooperation, the same holds for the number of listen predictions as cooperatively playing participants with a higher need for evidence would expect the same from their co-participants, because they know that their co-participants are also aiming to coordinate the actions of both participants.

Prediction accuracy. This is clearly the purest behavioral measure of the veracity of a participants' mental model of the other participant: if one participant can accurately predict the other participant's actions, then the mentalizing capacity is clearly successful. Prediction accuracy can be also seen as an expression of a successful Level 1 or higher ToM agent [20]. Such an agent is capable of representing what the other participant believes about the tiger's location and the other participant's expected values for actions.

Consistent actions. We define consistent actions as those participant choices that logically follow their predictions. For example, if a participant in a cooperative ITT believes that the other participant believes that the tiger is behind the left door and that the other participant will open the door on the right in the current round, the participant will predict open-right. If the participant also believes that the tiger is behind the left door, then the participant should also open the right door on this

round in order to maximize the probability of greatest reward. Of course, there will be rounds when a participant believes that the other participant is wrong about the location of the tiger, but such trials are expected to be infrequent. In the context of a cooperative ITT, where action coordination is beneficial for both participants, consistent actions are an important behavioral strategy for successful performance: if the participant predicts the other participant will choose a specific action, then he increases his own predictability, if he chooses an action that is consistent with his own predictions. As such consistent actions are a rudimentary indicator for Level 2 or higher ToM reasoning: if the participants in the cooperative ITT know that they need to exactly coordinate their actions, then it is beneficial for them to increase the predictability of their own actions. That is, assuming they are Level 2, they choose a strategy that maximizes the chances that they can be predicted by a Level 1 ToM agent, who is trying to predict their behavior as if they were a Level 0 agent (i.e., they have no ToM; they do not cognitively model another agent's beliefs or values). Of course, such sophistication of ToM reasoning is especially beneficial in the cooperative ITT and might prompt participants to engage in this "deeper" recursion of ToM reasoning (see also Discussion in section 4).

Correctly predicted consistent actions. Consistent actions are predicated on the participant's predictions of the other participant and these predictions can be erroneous. Therefore, we also look at a subset of consistent action, namely those in which the predictions of the other participant were correct.

Thinking about the other agent's potential choices and incorporating these predictions into one's own action selection process is cognitively demanding and requires considerable cognitive resources. We therefore expect an increase in reaction time (RT), whenever participants think about the other participant in detail. An important aspect of repeated social interaction in the ITT is that the participants can learn the other's decision strategies and choice preferences. We look at the learning-related changes that this index brings to the ITT behavior.

In a final step, we wanted to link the prediction performance of the participant indicative of their ToM process to their own choices and evaluate, whether successful predictions of the other participant also resulted in better choices. We, therefore, provide several correlation analyses in which we link our prediction measures with several of the choice indices from above.

2.6 Models of optimal performance

We aim to characterize the performance of our participants with respect to the performance of an optimal agent on the TT and the ITT. The computational model for the TT is a partially observable Markov decision process (POMDP), which has been introduced by Kaelbling et al. [12]. It is a generic framework for decision-making under uncertainty, when agent do not have direct knowledge to the state of the world, but only through probabilistic observations provided by the environment. To accommodate such a decision-making scenario in an interactive context, Gmytrasiewicz and Doshi developed interactive POMDPs (I-POMDPs) [13], a generalization of POMDPs to multiple interacting agent, which allows for the modeling of the other agents' beliefs and actions within the computational model itself. In the following, we briefly describe these two frameworks and how they can solve the TT and the ITT. For more details (including the equations defining the model) we refer the interested read to the original publications [12, 13].

2.6.1 POMDPs as a model for the single agent Tiger Task

In a POMDP an agent does not have direct knowledge of the state of the world. In the case of the TT the two possible states of the world are "Tiger Left" and "Tiger Right" indicating the door, which hides the tiger that should be avoided. However, each state emits probabilistic observations, which are presented to the agent. In the case of the TT, these are the physical observation of "Growl Left" and "Growl Right" indicating the location of the tiger with an accuracy of 70%. Using these observations in a Bayesian belief updating scheme, the agent forms beliefs about the world in form of a probability distribution over states specifying the probability of each state in the current trial. These beliefs, which are updated in a so-called *state estimator*, are the basis for the decision of the agent, which action to take next.

More formally, a POMDP is defined by:

- a set of state S, which define the environment
- a set of action A, which an agent can take in each state
- a state transition function T: S×A detailing the (possibly probabilistic) transitions between state of the environment conditional on the specific action
- a reward function R: S×A detailing the immediate reward that an agent obtains when selecting action a in state s
- an observation function O: S×A detailing the probability distribution of observations which the agent can make after performing action a in state s

The model performs Bayesian belief updating in the state estimator, which calculates the new beliefs of the agent based on the current state (unknown to the agent), the last action, and the observation that the agent has made. These updated beliefs (about the location of the Tiger) are the basis for the next action decision. These beliefs are filtered through softmax function (a sigmoid function) transforming beliefs into action probabilities, which are the model's predictions for the next action by the agent.

We used a the POMDP implementation pomdp-solve by Toni Cassandra (code available at `http://www.pomdp.org` to solve the TT using both the original and modified payout structure. The solver provides and optimal solution to POMDP in form of a policy graph that specifies the optimal action sequence given a specific observation and a particular belief. Specifically, we focused on the number of listen action that the optimal agent would take (given the observation sequence of the participants in each trial and compared this to the actual number of listen action of the participants (averaging both over trials). In the case when the optimal agent reached an open action before the participant, we took the preceding number of listen action as the optimal number for this trial; in the case, when the participant committed and open action before the optimal agent, we simulated 30 additional observation sequences using the observation probabilities defined in the TT and recorded the number of listen actions (averaging over these simulations).

2.6.2 I-POMDPs as a model for the multi-agent Interactive Tiger Task

An interactive POMDP (I-POMDP) generalizes the POMDP framework to a multi-agent setting, in which two or more agent simultaneously take actions in an uncertain environment. This is a new level of complexity, because now the action of the other player(s) have to be taken in account when

calculating the next action. To solve this problem in a optimal way, the I-POMDP creates a model of the other players, which calculates their beliefs and action probabilities and then uses these simulated beliefs to compute the agent's optimal action. Thus, an I-POMDP defines an "model within a model" and is thus a computational vehicle for assessing mentalizing in a quantitative way.

More formally, an I-POMDP has the same defining elements as a POMDP except that the set of state S is replaced by a set of interactive states IS: S×M, which is the interaction of the states with the possible models M of the other agent(s). These M models individually also contain the main agents' models which recursively do the same and so on. When calculating the beliefs, the I-POMDP first calculates the possible beliefs of the other agent(s) given the observations that the other agent(s) make, and then marginalizes over the other agents' beliefs to compute the own belief update. This is then filtered again through a quantal response equilibrium function to obtain action probabilities for the next decision. For more details on the I-POMDP and the implemented quantal response equilibrium function, please see [13, 25].

Including a model of the other agent in the belief update of the I-POMDP raises the question of the level of recursion of these "models within models" [26, 27]. A Level 0 I-POMDP agent (essentially a POMDP agent) does not have a model of the other agent and learns only from the observed environment. A Level 1 I-POMDP agent constructs a model the other agent as a Level 0 agent, so the first agent thinks that the other agent does not have a model of the first agent. A Level 2 agent builds a model of the other agent as a Level 1 agent, which includes a model of the first agent as a Level 0 agent. In summary, a Level n I-POMDP agent constructs a model of the other agent at the Level n-1. The level of recursion in the I-POMDP affects how the I-POMDP agent calculates and anticipates the other agent's actions.

Another important determinant of the belief update in the I-POMDP is the planning horizon. The value iteration algorithm that is used to solve the I-POMDP iterates over the number of planning steps while marginalizing over all possible own and other actions to calculate and update beliefs for the next decision. Especially for tasks like the ITT, in which there are several evidence-gathering steps, the valuation of actions, which are based on the beliefs, can be significantly affected by the number of steps the algorithm is allowed to look ahead while considering all possible joint actions.

For the comparison of the performance of our participants on the ITT with that of an optimal agent, we used a similar strategy than for the single-agent TT. However, due to high computational demand of the I-POMDP and optimal solution using value iteration [28] is commonly not possible. Hence, the optimal solution has to be approximated. We used an interactive particle filter implemented in C++ [29, 21] to approximate optimal performance given the actual observation sequences that our participants experienced. We also limit the optimal I-POMDP agent to be a Level 1 agent with a planning horizon of either 1 or 2 steps with 1000 particles.

As in the case of the single-agent TT, whenever the I-POMDP reached an open action before the participant in a particular trial we took this as the optimal number of listen actions for this trial; however, when the participant committed an open action before the I-POMDP, we simulated 30 possible observation sequences based on the observation probabilities of the ITT and recorded the number of listen action until the I-POMDP reached an open action. Then we average over simulations to obtain the optimal number of listen action for this trial. Finally, as in the case of the single-agent TT and the POMDP we averaged across all 30 tiger trials to obtain a subject-specific number of listen actions.

2.7 Statistical analyses

All the statistical tests were performed in R version 3.6.1. For comparisons of the means of the interactive contexts, we performed an independent Welch two-Sample t-test between the sample means and report the t-value, degrees of freedom, p-value, the 95% confidence interval and the effect size with Cohen's d-value. We also corrected for multiple comparisons using Bonferroni corrections, setting the criteria for the significance at the value of $p < 0.006$ (0.05/8). For the comparison between the optimal POMDP model and the participant behavior we performed a two-way mixed effects ANOVA treating participant behavior and the optimum POMDP model as the within-subject factor and the two matrix versions as the between-subject factor. Similarly, another two-way mixed effects ANOVA treating participant behavior and the optimum I-POMDP model as within-subject factor, and the two contexts as the between-subject factor was performed for the multiagent ITT.

For the comparisons of learning effect through sessions between contexts, we performed two-way mixed effects ANOVA treating session as a within-subject factor and context as a between-subject factor with significance set at $p < 0.05$. From the ANOVAs we report the F-statistic degree of freedom between and within, the F-ratio and the p-value. Differences in the correlation coefficients were tested at the significance level of $p < 0.05$ using Fishers' z-transformation by calculating the *z_observed* value.

$$z_{\text{observed}} = \frac{(z_1 - z_2)}{\sqrt{(1/N_1 - 3) + (1/N_2 - 3)}}$$

Where, z_1 and z_2 are the Fisher z-transforms of the two correlation coefficients, while N_1 and N_2 are their respective sample sizes. Statistical significance at the 5% level is achieved when this value is to be beyond a critical threshold of ± 1.96. A z_{observed} beyond the critical threshold is an indicator of significance and the rejection of null hypothesis.

3 Results

3.1 Single-agent version

We characterized the performance on the single-agent version of the tiger task (TT) on 3 indices: number of listen actions, evidence-difference, and percentage of correct open actions. (Please refer to section 2.5 describing the behavioral indices in more detail.) Because every participant completed the TT and this version only contains physical, or state, observations (tiger growls) without the complexities of social interactions and mentalizing, it provides the best context for studying the effects of the original and modified version of the payout matrix. Figure 5 shows the results.

We observed a difference ($t(90.96) = 3.45, p = 0.86 \times 10^{-3}$, 95% CI $[0.43, 1.58]$,Cohen $d = 0.56$) in the number of listen actions between the two different version of the TT. Participants in the modified version ($mean \pm s.d.$: 3.82 ± 1.31) required significantly fewer listen actions than in the original version (4.82 ± 2.16) before committing an open action. Dissecting the number of listen actions into the two different observations (tiger growl left or right) that the participants could make, we found a smaller evidence-difference in the modified (2.05 ± 0.55) than in the original TT (2.43 ± 0.80, $t(98.19) = 3.48, p = 0.76 \times 10^{-3}$, 95% CI $[0.16, 0.60]$,Cohen $d = 0.56$). This is likely reflective of the larger risk associated with the original TT due to a larger loss following an incorrect open action. Surprisingly, we also saw a significant difference pointing to a comparatively better performance in the modified version in terms of correct open actions ($t(158.89) = -2.57, p = 0.011$, 95% CI $[-0.59, 0.01]$,Cohen $d = 0.38$; See Figure 5).

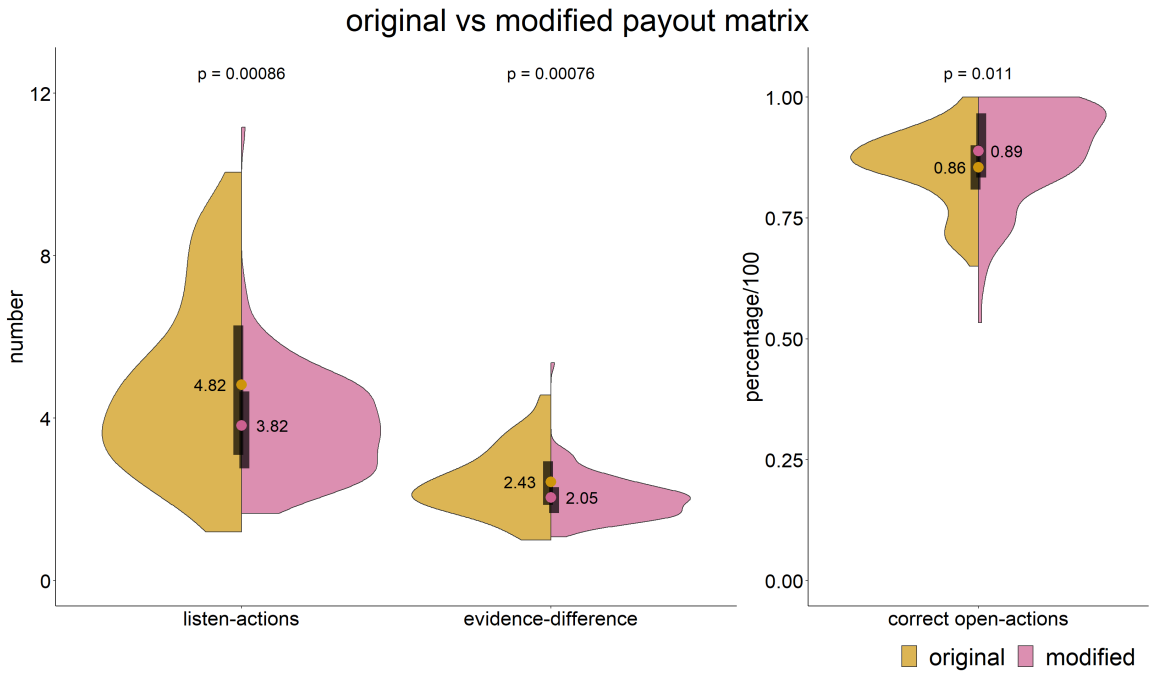

Figure 5: Behavioral indices comparison of the single-agent version of the TT. This Figure compares the participants' performance difference in the original version with the modified version of the payout matrix. It shows the mean values with their distributions in the form of the split violin plot. The black bars cover the interquartile range. The number of listen actions and the evidence-differences are significantly lower in the modified version. The number of correct open actions also was higher in the modified version, indicating the participants found it informationally easier or less risky to work with the modified version of the outcome matrix.

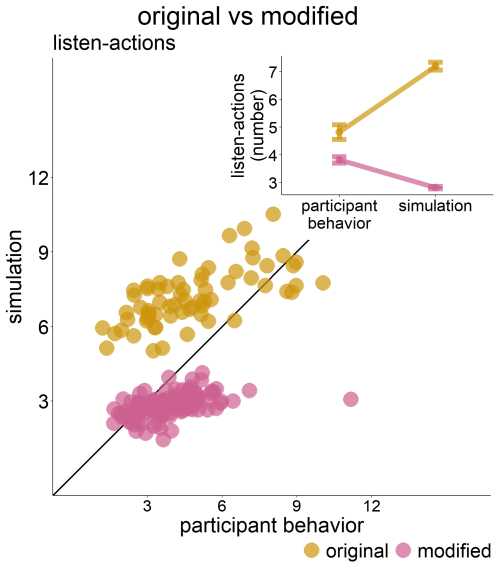

Figure 6: Comparison of actual and optimal number of listen action in the single-agent TT. Each dot represents a participant characterized by his actual (x-axis) and optimal (y-axis) number of listen actions averaged across all tiger trials. Whereas the optimal POMDP Level 1 model overestimates the number of listen action for the original payout matrix, is underestimates it for the modified payout matrix. In figure inset, we see main effects of the type of matrix, the participant behavior/simulation, and the interaction effect between them.

Did the performance of the participants deviate from the performance of the optimal POMDP agent?

We observed a stark difference between the original and the modified payout matrix in the TT (see Figure 6). For the original payout matrix the optimal POMDP solution consistently overestimated the participants' number of listen actions (mean nListen: POMDP 7.20 ± 1.14, participants 4.82 ± 2.16, $t(98.29) = -7.91, p = 3.97 \times 10^{-12}$, 95% CI $[-2.98, -1.78]$,Cohen $d = 0.97$). However, for the modified payout matrix the POMDP underestimated the participants' actual number of listen actions (mean nListen: POMDP 2.82 ± 0.46, participants 3.82 ± 1.31, $t(152.71) = 8.01, p = 2.67 \times 10^{-13}$, 95% CI $[0.75, 1.24]$,Cohen $d = 0.72$). Using a 2-way mixed-effects ANOVA with the within-subject factor as participant behavior and optimal POMDP, and the between-subject factor as the original and modified matrix, we observed the main effect of the type of matrix ($F(1, 374) = 308.41, p < 2.00 \times 10^{-16}$), the participant behavior and simulation ($F(1, 374) = 24.33, p < 1.22 \times 10^{-6}$), and also an interaction effect between them ($F(1, 374) = 121.76, p < 2.00 \times 10^{-16}$) (see Figure 6 inset).

3.2 Multiagent version

In the main text of this paper we focus on the results with the modified payout matrix of the multiagent ITT. All subsequent findings in this section therefore, refer to the modified ITT. However, we provide results from the original payout matrix in the supplement as these findings may be of interest for the wider community using the published version of the multiagent Tiger Problem for model development and simulations.

3.2.1 Interactive effects on choice indices

We first examined the effect of the interactive context on the choice indices defined above in section 2.5. We observed a significant effect on number of listen actions ($t(102.49) = -4.54, p = 1.6 \times 10^{-5}$, 95% CI $[-2.76, -1.08]$,Cohen $d = 0.82$), which was higher in cooperation ($mean \pm s.d.$: 5.16 ± 1.93) than in competition (3.23 ± 2.67) (see Figure 7). This suggests that cooperative participants engaged in prolonged evidence gathering in forming their estimates of the tiger location. Comparing the number of listen actions to the single agent TT, we saw a significant difference in ITT cooperation ($t(97.59) = -5.04, p = 2.155 \times 10^{-6}$, 95% CI $[-1.86, -0.81]$,Cohen $d = 0.57$), while not in ITT competition ($t(70.13) = 1.58, p = 0.12$, 95% CI $[-0.15, 1.32]$,Cohen $d = 0.20$). Under cooperation, participants knew that maximal reward was achieved by coordinated open actions, and so they would not have had incentive to race one another to the door. Conversely, under competition, participants were expected to view the task as a race so as to beat the other participant to the correct door and receive maximal reward. That is, participants would have likely underestimated their own probability of opening the wrong door by overestimating the weight of loss if their competitor would have opened the correct door.

This was paralleled by a significant effect on evidence difference between interactive contexts. In the competitive context, the participants preferred to open the door with less evidence (1.56 ± 0.57) than in the cooperative context (2.32 ± 0.60) ($t(121.14) = -7.12, p = 8.2 \times 10^{-11}$, 95% CI $[-0.96, -0.54]$,Cohen $d = 1.28$), pointing towards increased coordination during cooperation and increased risk-taking during competition. These difference were significantly less when compared to the TT ($t(58.83) = -3.80, p = 3.5 \times 10^{-4}$, 95% CI $[-2.05, -0.63]$,Cohen $d = 0.61$) and more ($t(69.09) = -12.49, p < 2.2 \times 10^{-16}$, 95% CI $[-3.50, -2.53]$,Cohen $d = 0.33$) for the competitive and cooperative context respectively.

We further observed a higher percentage of identical open actions in the cooperative (0.40 ± 0.20) compared to the competitive context (0.32 ± 0.13, $t(113.36) = -2.86, p = 4.9 \times 10^{-3}$, 95% CI

$[-0.15, -0.03]$,Cohen $d = 0.51$) suggesting that participants in the cooperative context, which incentivized action coordination, increased successful coordination.

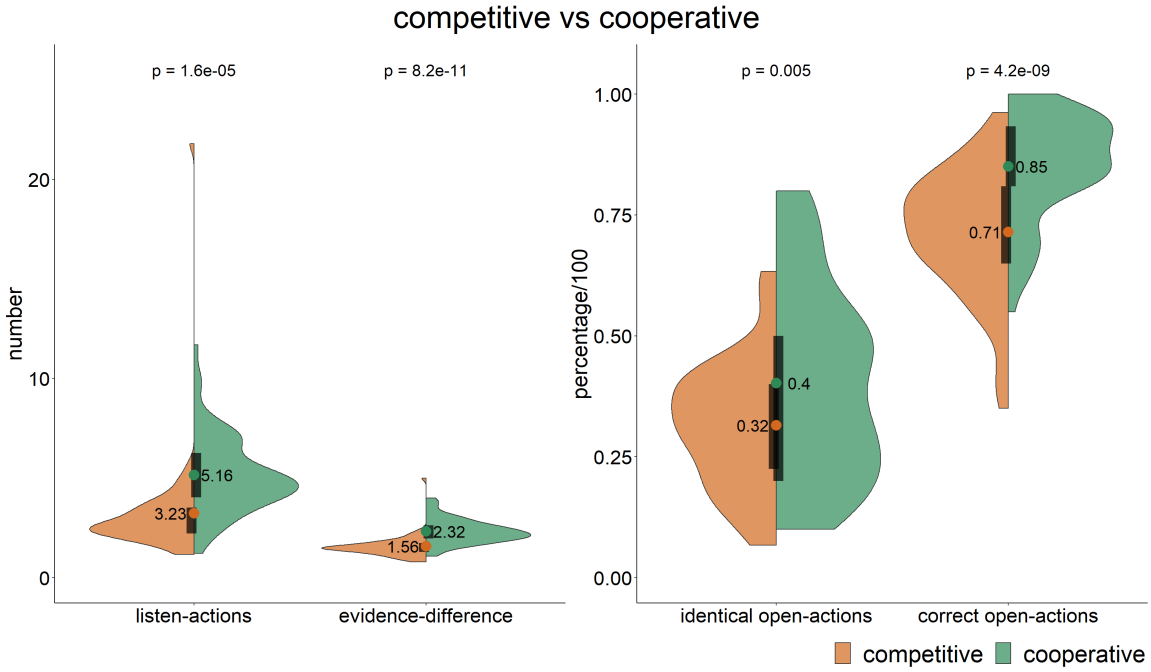

Figure 7: Interactive choice behavioral indices' comparison of the multiagent version of the ITT (modified payout only), competitive and cooperative contexts. The figure shows the mean values with their distributions in the form of split violin plots. The black bars cover the interquartile range. The number of listen actions and the evidence-difference are significantly higher in the cooperative context, indicating greater evidence gathering for coordination in the task. The percentage of identical open actions was also significantly higher in the cooperative context which translated in the higher correct open actions as well.

Finally, we also found a significant difference in the percentage of correct open actions, which was again higher during cooperation (0.85 ± 0.11) than competition (0.71 ± 0.13, $t(111.55) = -6.38, p = 4.2 \times 10^{-9}$, 95% CI $[-0.18, -0.09]$,Cohen $d = 1.15$). This suggests that participants in the cooperative context might have been able to utilize their longer evidence gathering (number of listen actions) effectively for more successful open actions.

How did the performance of our participants deviate from optimal performance of an I-POMDP agent? To answer this question we again focused on the number of listen actions and compared participant performance to an I-POMDP agent with a planning horizon of 1 and 2 steps.

The Level 1, Horizon 1 (L1H1) I-POMDP with a 1000 particles was not able to discriminate between different number of listen action in the cooperative and the competitive context. It always predicted an optimal number of listen action to be around 1.5 regardless of the interactive context (see Figure 8A) with an initial belief of tiger being behind either of the doors at 50%. In contrast, the Level 1, horizon 2 (L1H2) I-POMDP agent over- and underestimated the actual number of listen actions of the participants in a particular way: in the cooperative context it underestimated the participants number of listen action (mean nListen: I-POMDP, 3.83 ± 0.39, participant 5.16 ± 1.93, $t(70.37) = 5.46, p = 6.8 \times 10^{-7}$, 95% CI $[0.84, 1.81]$,Cohen $d = 0.67$), whereas in the competitive context, it overestimated the participants' number of listen actions (mean nListen: I-POMDP, 5.53 ± 0.45, participant 3.23 ± 2.67, $t(60.27) = -6.46, p = 2.0 \times 10^{-8}$, 95% CI $[-3.01, -1.59]$,Cohen $d = 0.85$) (see

Figure 8B). Using a 2-way mixed-effects ANOVA with the within-subject factor as participant behavior and optimal L1H2 I-POMDP, and the between-subject factor as context, we observed an interaction effect ($F(1, 242) = 72.72, p < 1.64 \times 10^{-15}$) (see Figure 8B inset). This suggests a distinctly different pattern in the optimal I-POMDP agent than in the participants: the optimal agent would listen more in the competitive context compared to the cooperative context, whereas for the participants this pattern is reversed.

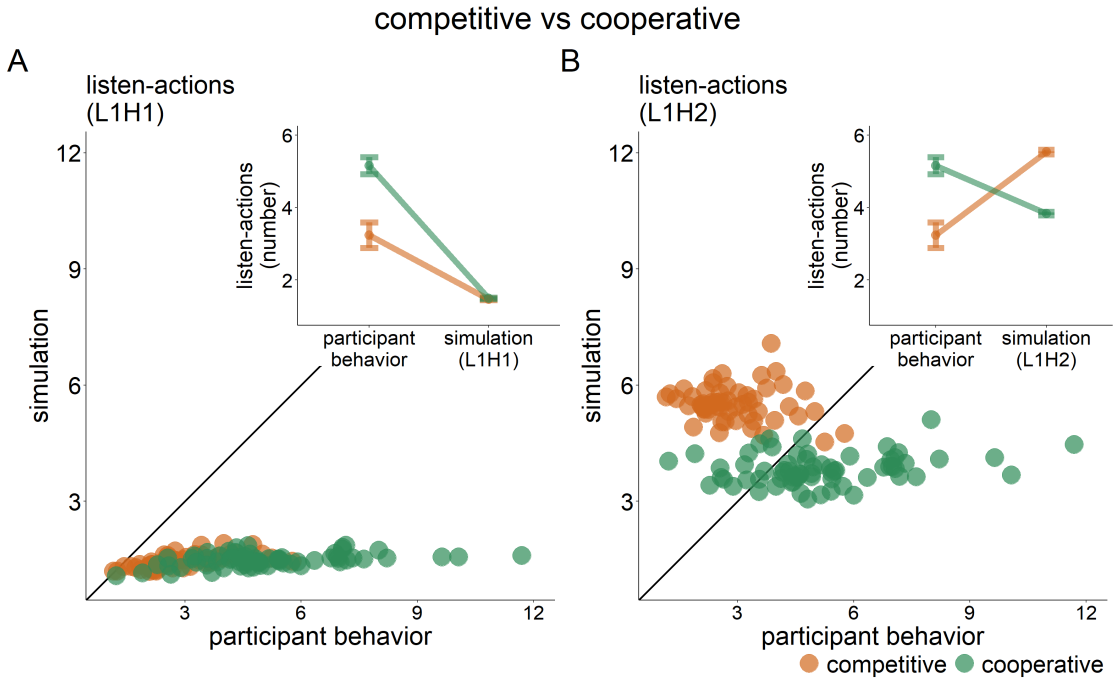

Figure 8: Comparison of actual and optimal number of listen action in the multiagent ITT. Each dot represents a participant characterized by his actual (x-axis) and optimal (y-axis) number of listen actions averaged across all tiger trials. [A] For L1H1, the optimal I-POMDP underestimates the number of listen actions in both the contexts. The figure inset shows the interaction effects of the context. [B] While for the L1H2, the optimal I-POMDP overestimates the number of listen action in the competitive context and it slightly underestimates it in the cooperative context. The figure inset in [B] shows the context x data type interaction effect.

3.2.2 Interactive effects on prediction indices

In the next step we examined the effects of the interactive context on behavioral indices of predictions (see Figure 9). These analyses also shed some light on the cognitive processes involved in modeling the beliefs and choices of the other participant. We first compared prediction accuracies in both interactive contexts. Participants exhibited more accurate predictions of the other participant in the cooperative context (0.91 ± 0.45) compared to the competitive context (0.81 ± 0.08, $t(89.47) = -8.82, p = 8.6 \times 10^{-14}$, 95% CI $[-0.13, -0.08]$,Cohen $d = 1.61$), indicating that participants responded to the demand characteristics of the cooperative ITT, which incentivizes precise coordination of actions and hence a demand for accurate predictions of the other participant's actions.

Paralleling the higher number of listen actions in the cooperative ITT, we also observed a higher number of listen predictions during cooperation (5.03 ± 1.87) than during competition ($3.00 \pm 2.34, t(109.04) = 5.29, p = 6.42 \times 10^{-7}$, 95% CI $[-2.80, -1.27]$,Cohen $d = 0.96$). As cooperating participants demonstrated longer evidence gathering before an open action, they may have projected the same tendency onto their co-participants, leading to a higher number of listen predictions.

Similarly, we also found a higher percentage of consistent actions during cooperation (0.98 ± 0.03) than competition (0.90 ± 0.08, $t(67.79) = -7.50, p = 1.8 \times 10^{-10}$, 95% CI $[-0.10, -0.06]$,Cohen $d = 1.38$). This is consistent with an expectation that participants would act more predictably during cooperation than competition. As explained in Section 2.5, this can be interpreted as rudimentary for sophisticated Level 2 ToM reasoning, as the cooperating participants may take the perspective of a Level 1 ToM participant and act in a way that would help the Level 1 ToM participant be successful at coordinating their actions with the participant. Moreover, we also observed that among the consistent actions, those that were correctly predicted for the other participant were higher during cooperation (0.90 ± 0.05) than competition (0.75 ± 0.10, $t(80.77) = -10.06, p = 6.8 \times 10^{-16}$, 95% CI $[-0.18, -0.12]$,Cohen $d = 1.85$).

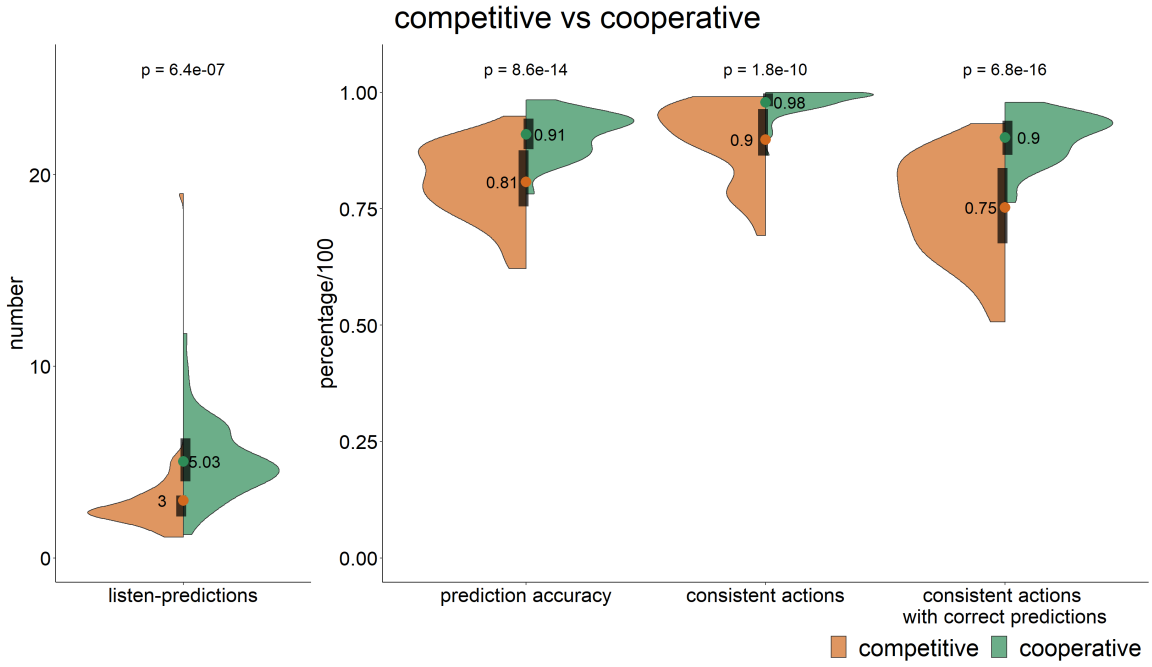

Figure 9: Interactive prediction behavioral indices: comparison of the competitive and cooperative ITT. Mean values with their distributions in the form of split violin plots. The black bars cover the interquartile range. The number of listen predictions are significantly higher in the cooperative context. The higher prediction accuracy in the cooperative ITT indicates that the participants were better able to anticipate the others' choice behavior. Consistent actions and consistent actions following correct predictions of others' choices are important criteria in achieving favorable outcomes in the cooperative context, as seen with significantly higher percentages during cooperation.

This line of interpretation is supported by an analysis of RTs for consistent and non-consistent action in both interactive contexts. After log-transforming the RTs to approximate a Gaussian distribution, we used a 2-way mixed-effects ANOVA with the within-subject factor action consistency and the between-subject factor context. We did not observe any main ($F(1, 115) = 0.78, p = 0.78$) or interaction effect ($F(1, 115) = 3.24, p = 0.074$) (see Figure 10A) with higher RTs for non-consistent actions in the cooperative ITT (6.68 ± 0.36) compared to consistent actions (6.60 ± 0.26). RTs for both types of actions were comparable in the competitive ITT (consistent actions: 6.64 ± 0.25, non-consistent actions 6.60 ± 0.26).

Subsequently, we sub-divided RTs for consistent actions into those associated with correct and incorrect predictions in both interactive contexts. A 2-way mixed-effects ANOVA showed a main effect

of prediction accuracy (correctly predicted consistent actions: 6.64 ± 0.23, incorrectly predicted consistent actions: 6.59 ± 0.25, $F(1, 122) = 15.42, p = 0.143 \times 10^{-3}$), but no main effect of context ($F(1, 122) = 0.29, p = 0.592$) and no interaction ($F(1, 122) = 0.02, p = 0.888$) (see Figure 10B). This is consistent with the notion that accurate mentalizing about the other participant, which likely leads to accurate predictions, requires more cognitive resources, leading to longer RTs regardless of the interactive context.

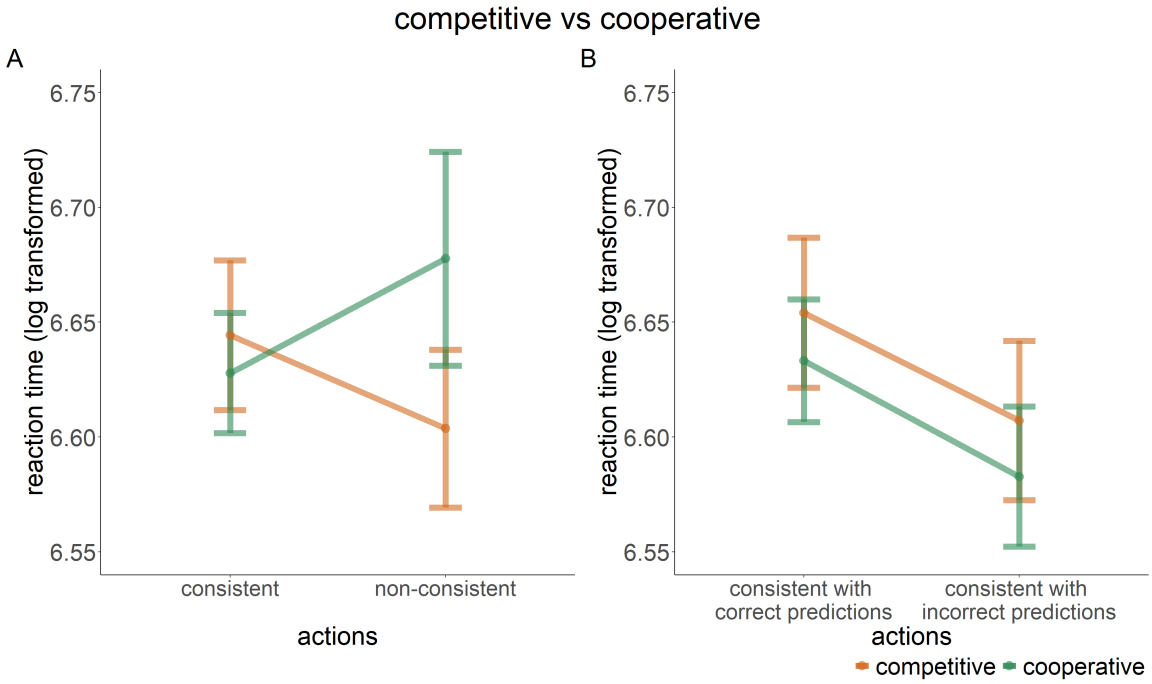

Figure 10: Comparison of RTs. [A] RTs for consistent and non-consistent actions. Using 2-way mixed effects ANOVA we observed no interaction effect of the context and the consistency of actions. [B] RTs for correctly and incorrectly predicted consistent actions. Using 2-way mixed effects ANOVA we observed a main effect of accuracy, and no interaction effects. All error bars show mean +/- the standard error.

3.2.3 Relating choice and prediction indices

To answer the question of whether better prediction accuracy resulted in improved dyadic coordination and in improved overall performance, we related the differences in prediction performance in the two contexts to the choice performance of our participants (Figure 11) We observed a high correlation between the prediction accuracy and the identical actions in all participants (cooperative: $r = 0.90$, 95% CI $[0.84, 0.94]$; competitive: $r = 0.76$, 95% CI $[0.62, 0.85]$) suggesting that better prediction performance improved the coordination of actions. The correlation coefficients are significantly different ($z_{\text{observed}} = -2.55, p < 0.05$) (Figure 11A). We also observed significant correlation between prediction accuracy and identical open actions (cooperative: $r = 0.61$, 95% CI $[0.43, 0.74]$; competitive: $r = 0.46$, 95% CI $[0.22, 0.64]$), but no difference ($z_{\text{observed}} = -1.14$) due to ITT condition (Figure 11B).

To investigate the link between listen actions and listen predictions we correlated the number of listen actions and the number of listen predictions, finding associations in both ITT conditions (cooperative: $r = 1.00$, 95% CI $[0.99, 1.00]$; competitive: $r = 0.99$, 95% CI $[0.99, 1.00]$), and found no difference between them ($z_{\text{observed}} = -1.33$). However, we saw greater individual variability in the competitive context than in the cooperative context, indicating a higher mismatch between the participants predicting the other participants' action choices (Figure 11C). Taken together these analyses are consistent

with the expectation that prediction and choice performance are closely related in both interactive contexts, but that cooperation elicits a stronger relationship between prediction and choice.

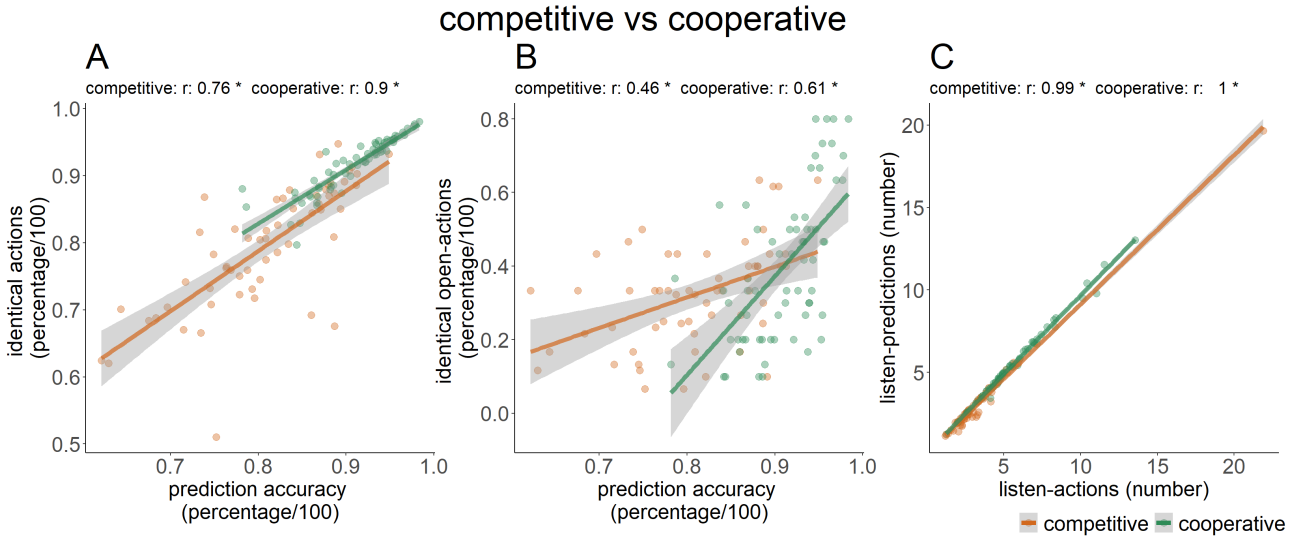

Figure 11: Correlations relating choice and prediction indices. We observed significant correlation between [A] prediction accuracy and identical actions with significance between the contexts. We also observed significant correlation between [B] prediction accuracy and identical open actions and between [C] the number of listen actions and number of listen predictions.

3.2.4 Learning-related improvement in task performance

We expected that participants would learn about the other participants' preferences and beliefs through repeated interactions in the ITT, so we should therefore observe a change in these indices across all three sessions. Figure 12 shows that most learning occurred between sessions 1 and 2.

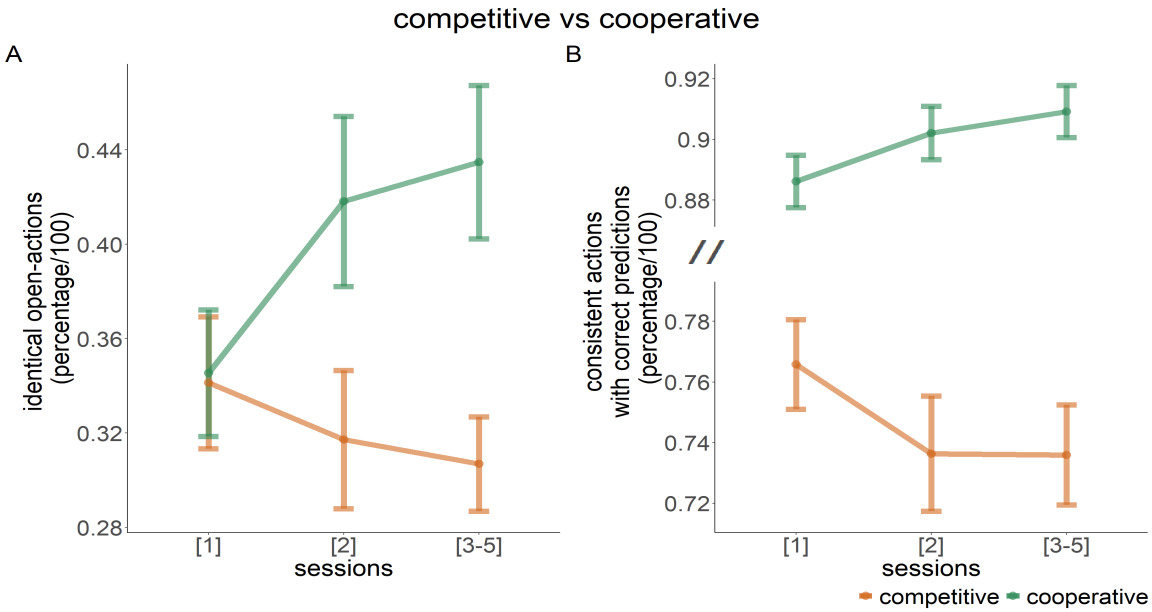

Figure 12: Learning in the ITT. [A] Identical open actions and [B] correctly predicted consistent actions over the course of sessions. Using 2-way mixed effects ANOVA, in [A] we observed a main effect of context and a significant interaction indicating that participants in the cooperative context learn to simultaneously open the door through subsequent sessions, compared to the competitive context. In [B], for the correctly predicted consistent actions we see a similar effect of context and session. Error bars are s.e.m.

Paralleling our findings on identical actions in Figure 7, we saw a main effect of cooperation vs. competition on identical open actions ($F(1, 122) = 6.04, p = 0.015$), with higher numbers of identical open actions during cooperation, increasing over sessions (interaction effect context x session $F(2, 244) = 3.69, p = 0.026$).

The number of consistent actions with correct predictions also diverged between contexts over the course of the sessions (interaction effect context x session $F(2, 244) = 4.16, p = 0.017$). While this index increased slightly during cooperation, it substantially decreased during competition $F(1, 122) = 115.2, p < 2 \times 10^{-16}$. This is consistent with the expected demand characteristics of the two versions of the ITT. During cooperation the participants tried to be as predictable as possible to ensure the best chances for coordinated actions, while during competition they aimed to be unpredictable to ensure that they would not lose to the opponent.

4 Discussion

We adapted the single- and multiagent Tiger Problem to create a single-agent Tiger Task (TT) and multiagent, interactive Tiger Tasks (ITT), to better understand how human agents depart from computationally optimal solutions under conditions of uncertainty and partial observation. We have provided the first empirical report of data from human participants on single and multiagent versions of adaptations of the Tiger Problem. The Tiger Problem is an iconic challenge in the artificial intelligence community as it seeks to develop sophisticated models for planning under uncertainty specifically for adaptive interactions with other agents [12, 21]. These models (POMDP/I-POMDP) capture the belief updating process, in which the agent has only partial access to information about the current state that is relevant to probabilistic rewards, manipulating the factors of both interactivity and uncertainty [20].

Along with assessing how human behavior departs from that consistent with normative models, the goal of this paper was to characterize the empirical choices of human participants and their predictions of the other agents' actions, using only model-free behavioral indices. These indices revealed a number of important differences (discussed below) between the performance of the single agent and multiagent task and between the cooperative and the competitive ITT. Importantly, these empirical, model-free findings can serve as targets for posterior predictive checks for a more detailed modeling of beliefs and valuations in subsequent work. Also we will further explore the neural underpinnings of the interactive social decision making underpinning behavior in our newly designed tasks [30, 31].

Participants completed one of two single-agent versions of the TT that differed in the size of the outcomes as prescribed by the fixed payoff matrices. Subsequently, they engaged in either a cooperative or a competitive version of the ITT, which differed in the relative size of the outcome given specific joint actions. The cooperative ITT favored strictly coordinated actions (i.e. identical actions), whereas the competitive version of the ITT favored making correct open actions before the other participant, and yielded the greatest reward when an agent opened the correct door as the other agent opened the door to the tiger.

In the TT, we found that reduced risk of loss in the modified payout matrices increased how well the optimal model matched the number of listen actions from human agents (Figure 6). This is consistent with our hypothesis that participants underestimated their probability of loss while overestimating the weight of the small losses incurred by listening [32, 33].

As expected we found higher scores of the cooperative participants on most of our indices of choice and prediction performance in the ITT (Figures 7 and 9). Cooperation elicited more listen actions and more predictions of listen actions. These findings associated with higher prediction accuracy and a higher proportion of identical actions and correct open actions in cooperation compared to competition. These results suggest that participants adopted successful strategies more often in the cooperative version of the ITT. Strict coordination of action in the cooperative context, especially of open actions, requires accurately assessing when one's own belief about the location of the tiger matches the belief of one's co-participant. This would be achieved by prolonged evidence gathering, leading to an increased evidence differences observed in the cooperative ITT, which is a measure of the consistency of the evidence of the tiger location. Likely due to the longer evidence-gathering phase, participants in the cooperative ITT were more successful at opening the correct door. In contrast, participants playing the competitive ITT were characterized by a shortened evidence-gathering phase, consistent with the notion that they raced one other to the door, which comes at the price of a smaller number of correct open actions. These participants also showed evidence of learning to avoid identical open actions.

Comparing human behavior with simulations of optimal performance as generated from an I-POMDP reveals several interesting differences in the number of listen actions per tiger trial. While the participants committed more listen actions in the cooperative context (Figure 7) than under competition, horizon 1 models (L1H1) predict no such difference. This suggests that human participants did not simply rely on information in the current round and perhaps used deeper levels of recursive thinking to guide their actions. Modeling will allow us to test these possibilities. On the other hand, the horizon 2 models (L1H2) show that it is optimal to listen more under competition and less under cooperation, which is the opposite pattern seen for human participants (Figure 8B, inset).

There are two possible explanations for this difference. One proposal says that participants did not engage with the other participant in their dyads at all and only paid close attention to their own marginal utilities in the payoff matrices. Figure 4 shows that a listen action under competition has a positive marginal utility of +14 points in the modified payout matrix, whereas under cooperation a listen action has a marginal utility of -16 points. Competition also yields a much larger marginal utility for correct open actions (+95) compared to cooperation (+15 points). The differential for listen actions increases the value of listen actions during competition, which should lead to longer stretches of listen actions under competition, as predicted by the L1H2 model. If human agents departed from optimality by attending only to gains and not at all to losses, however, and if they ignored the presence of the other participant in their dyads entirely, the marginal utilities of listen and correct open actions would drive them to listen less under competition than under cooperation. Such an interpretation would suggest biased processing of reward information in humans, such that loss probabilities would be discounted completely. This result would challenge Prospect Theory, one of the most prevalent theoretical accounts of human decision-making [34]. This proposal predicts that participants' actions would be best modeled by eliminating ToM and remaining with L0 models, given the inattention to dyadic partners and exclusive attention to marginal utilities of listen actions and correct open actions.

An alternative proposal is consistent with the notion that the participants did attend to and take account of the presence of dyadic partners during interaction. In this proposal, competition leads participants to overestimate the probability that the dyadic partner will reach the correct door first

and so to overestimate the probability of the associated larger losses, while also overestimating the small losses incurred by listening. Overweighting losses and underweighting gains is consistent with Prospect Theory. Additionally, this alternative proposal predicts that participants' actions are not well modeled unless one uses at least L1 ToM, given the attention to dyadic partners. Future modeling with I-POMDP models will allow direct assessment of whether L0, L1, or deeper levels of recursive ToM best account for participants' actions.

Examining the data regarding predictions provides an indirect or oblique assessment of possible ToM processes. As expected, cooperation elicited higher prediction accuracy than did competition, which was likely related both to the prolonged evidence-gathering phase during cooperation and to increased consistency of action among participants under cooperation. The higher number of listen predictions in the cooperatively playing participants also associated with increased action coordination.

Asking participants to predict the other participants' actions – a novel task element that was not part of the multiagent Tiger Problem – was intended to elicit explicit ToM processes. Increased prediction accuracy is an indirect measure of the success of this putative cognitive activity. It can be interpreted as the first step towards the recursive ToM thought process, where "I think that you think..." becomes "I think that you think that I think..." and so on. Although cooperation led to greater accuracy in predicting other participants' actions, this is likely to be primarily due to prolonged evidence-gathering via listen actions, and not primarily due to cooperation facilitating better social cognition via ToM per se. However, our future work will directly test whether cooperation does facilitate better ToM, in addition to leading to greater evidence gathering.

An interesting aspect of the predictions are the consistent actions, in which participants chose the same action that they predicted for the other participant. Consistent with our hypotheses, cooperation elicited more consistent actions than competition. This can be interpreted as an effect of the incentives for coordinated actions in the cooperative payout matrix, but it could reveal an interesting further step in a recursive hierarchy of ToM processes. Through consistent actions, a participant in a cooperative dyad increases their own predictability for the other participant, so as to maximize the chances that they will end up choosing incentivized identical actions. This may indicate that participants adopt the perspective of the other participant and think about what action they themselves should take to make it easier for the other participant to choose the same action. However, the mere fact that cooperatively playing participants chose more consistent actions does not definitively indicate a higher level of ToM processes. Our future work will explore this possibility in much greater detail using computational models like I-POMDPs that can build "a model within a model" of the other participant.

A further interesting aspect of the data is that during cooperation, participants exhibited shorter reaction times for consistent compared to non-consistent actions. This would be expected if consistent actions require less cognitive effort, since they are aligned with their strategic goals in the cooperative ITT (namely to be predictable in their action selection). Non-consistent actions violate these goals, which may be the reason why it takes more effort and time to make them. However, the RT analyses also revealed that among the consistent actions, people took longer when correctly predicting actions of others. This suggests that successful reasoning about other participants' actions requires more cognitive resources and thus more time. There has been previous work exploring this in more detail [35].

Participants learned during the ITT sessions. Namely, identical actions and correctly predicted consistent actions increased during cooperation but decreased during competition. These learning-related

changes suggest that participants in both interactive contexts learned to adopt the incentives implied in the respective payout matrices. During cooperation participants learned to coordinate their actions and improve their own predictability as they learned to adapt to their co-participants, whereas during competition, participants learned to become less predictable.

In conclusion, our study is the first to provide empirical data from human participants as they engage in adaptations of the single-agent and cooperative and competitive versions of the multiagent, interactive Tiger Problem. Human agents departed quite distinctly from computationally optimal choices. These deviations could have resulted from biased weights on losses and gains and/or on how competition and cooperation differentially bias representations and expectations of others. Future attempts to move from computationally optimal models of single agent and multiagent decision making [8] to applications with human persons and groups should proceed with caution and with empirical behavioral data. Otherwise, deployment of these technologies in real situations with real human beings will be less successful than is possible or will fail completely.

5 Acknowledgements

We thank the contributions of Julia Spilcke-Liss, Julia Majewski, Freya Leggemann, Franziska Sikorski and Vivien Breckwoldt with the large amount of data collection. We thank Shannon Klotz, Corinne Donnay, and Rena Patel for help with piloting and initial data assessment. SSK, PD, MS and JG were funded by a Collaborative Research in Computational Neuroscience grant awarded jointly by the German Ministry of Education and Research (BMBF, 01GQ1603) and the United States National Science Foundation (NSF, 1608278). JG and TR were supported by the Collaborative Research Center TRR 169 "Crossmodal Learning" funded by the German Research Foundation (DFG) and the National Science Foundation of China (NSFC). MS gratefully acknowledges support from a Scripps College Faculty Research grant. All authors declare no conflict of interest.

References

[1] Shomik Jain et al. "Modeling Engagement in Long-Term, In-Home Socially Assistive Robot Interventions for Children with Autism Spectrum Disorders". In: *Science Robotics* 5.39 (Feb. 26, 2020), eaaz3791. ISSN: 2470-9476. DOI: 10.1126/scirobotics.aaz3791. arXiv: 2002.02453. URL: http://arxiv.org/abs/2002.02453.

[2] Chris Birmingham et al. "Can I Trust You? A User Study of Robot Mediation of a Support Group". In: *arXiv:2002.04671 [cs]* (Feb. 11, 2020). arXiv: 2002.04671. URL: http://arxiv.org/abs/2002.04671.

[3] Min Hun Lee et al. "Designing Personalized Interaction of a Socially Assistive Robot for Stroke Rehabilitation Therapy". In: *arXiv:2007.06473 [cs]* (July 13, 2020). arXiv: 2007.06473. URL: http://arxiv.org/abs/2007.06473.

[4] Xiaobu Yuan. "Collaborative planning of assembly sequences with joint intelligence". In: *2011 IEEE International Conference on Robotics and Automation*. 2011 IEEE International Conference on Robotics and Automation (ICRA). Shanghai, China: IEEE, May 2011, pp. 134–140. ISBN: 978-1-61284-386-5. DOI: 10.1109/ICRA.2011.5979680. URL: http://ieeexplore.ieee.org/document/5979680/.

[5] Dharmashankar Subramanian et al. “A cognitive assistant for risk identification and modeling”. In: *2017 IEEE International Conference on Big Data (Big Data)*. 2017 IEEE International Conference on Big Data (Big Data). Boston, MA: IEEE, Dec. 2017, pp. 1570–1579. ISBN: 978-1-5386-2715-0. DOI: 10.1109/BigData.2017.8258091. URL: http://ieeexplore.ieee.org/document/8258091/.

[6] Nina Grgić-Hlača, Christoph Engel, and Krishna P. Gummadi. “Human Decision Making with Machine Assistance: An Experiment on Bailing and Jailing”. In: *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW Nov. 7, 2019), pp. 1–25. ISSN: 2573-0142, 2573-0142. DOI: 10.1145/3359280. URL: https://dl.acm.org/doi/10.1145/3359280.

[7] Vinay Kulkarni et al. “A Wide-Spectrum Approach to Modelling and Analysis of Organisation for Machine-Assisted Decision-Making”. In: *Enterprise and Organizational Modeling and Simulation*. Ed. by Joseph Barjis, Robert Pergl, and Eduard Babkin. Vol. 231. Series Title: Lecture Notes in Business Information Processing. Cham: Springer International Publishing, 2015, pp. 87–101. ISBN: 978-3-319-24626-0. DOI: 10.1007/978-3-319-24626-0_7. URL: http://link.springer.com/10.1007/978-3-319-24626-0_7.

[8] Marie-Pierre Pacaux-Lemoine and Frank Flemisch. “Layers of shared and cooperative control, assistance, and automation”. In: *Cognition, Technology & Work* 21.4 (Nov. 2019), pp. 579–591. ISSN: 1435-5558, 1435-5566. DOI: 10.1007/s10111-018-0537-4. URL: http://link.springer.com/10.1007/s10111-018-0537-4.

[9] A. Tversky and D. Kahneman. “The framing of decisions and the psychology of choice”. In: *Science* 211.4481 (Jan. 30, 1981), pp. 453–8. URL: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7455683.

[10] D. Talmi and A. Pine. “How costs influence decision values for mixed outcomes”. In: *Front Neurosci* 6 (2012), p. 146. ISSN: 1662-453X (Electronic) 1662-453X (Linking). DOI: 10.3389/fnins.2012.00146. URL: https://www.ncbi.nlm.nih.gov/pubmed/23112758.

[11] Andreea O. Diaconescu et al. “Inferring on the Intentions of Others by Hierarchical Bayesian Learning”. In: *PLOS Computational Biology* 10.9 (Sept. 4, 2014). Publisher: Public Library of Science, e1003810. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003810. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003810.

[12] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. “Planning and acting in partially observable stochastic domains”. In: *Artificial Intelligence* 101.1 (May 1998), pp. 99–134. ISSN: 00043702. DOI: 10.1016/S0004-3702(98)00023-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S000437029800023X.

[13] Piotr J. Gmytrasiewicz and Prashant Doshi. “A framework for sequential planning in multi-agent settings”. In: *Journal of Artificial Intelligence Research* 24 (2005), pp. 49–79. ISSN: 10769757. DOI: 10.1613/jair.1579.

[14] E. N., John von Neumann, and Oskar Morgenstern. “Theory of Games and Economic Behavior”. In: *The Journal of Philosophy* 42.20 (Sept. 27, 1945), p. 550. ISSN: 0022362X. DOI: 10.2307/2019327. URL: http://www.pdcnet.org/oom/service?url_ver=Z39.88-2004&rft_val_fmt=&rft.imuse_id=jphil_1945_0042_0020_0550_0554&svc_id=info:www.pdcnet.org/collection.

[15] George E. Monahan. “State of the Art—A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms”. In: *Management Science* 28.1 (Jan. 1982), pp. 1–16. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.28.1.1. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.28.1.1.

[16] Kaelbling, Leslie Pack, Littman, Michael L., and Moore, Andrew W. "Reinforcement Learning: A Survey". In: *Journal of Artifcial Intelligence Research* 4 (1996), pp. 237–285.

[17] Richard D. Smallwood, Edward J. Sondik, and Fred L. Offensend. "Toward an Integrated Methodology for the Analysis of Health-Care Systems". In: *Operations Research* 19.6 (Oct. 1971), pp. 1300–1322. ISSN: 0030-364X, 1526-5463. DOI: 10.1287/opre.19.6.1300. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.19.6.1300.

[18] Richard D. Smallwood and Edward J. Sondik. "The Optimal Control of Partially Observable Markov Processes over a Finite Horizon". In: *Operations Research* 21.5 (Oct. 1973), pp. 1071–1088. ISSN: 0030-364X, 1526-5463. DOI: 10.1287/opre.21.5.1071. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.21.5.1071.

[19] James N. Eagle. "The Optimal Search for a Moving Target When the Search Path Is Constrained". In: *Operations Research* 32.5 (Oct. 1984), pp. 1107–1115. ISSN: 0030-364X, 1526-5463. DOI: 10.1287/opre.32.5.1107. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.32.5.1107.

[20] Tessa Rusch et al. "Theory of mind and decision science: Towards a typology of tasks and computational models". In: *Neuropsychologia* 146 (Sept. 2020), p. 107488. ISSN: 00283932. DOI: 10.1016/j.neuropsychologia.2020.107488. URL: https://linkinghub.elsevier.com/retrieve/pii/S0028393220301597.

[21] Prashant J Doshi. "Optimal Sequential Planning in Partially Observable Multiagent Settings". PhD thesis. Chicago, Illinois: University of Illinois, 2005.

[22] Daniel C. Dennett. "Beliefs about beliefs [P&W, SR&B]". In: *Behav Brain Sci* 1.4 (Dec. 1978), pp. 568–570. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X00076664. URL: https://www.cambridge.org/core/product/identifier/S0140525X00076664/type/journal_article.

[23] Penelope A. Lewis et al. "Higher order intentionality tasks are cognitively more demanding". In: *Social Cognitive and Affective Neuroscience* 12.7 (July 1, 2017), pp. 1063–1071. ISSN: 1749-5016, 1749-5024. DOI: 10.1093/scan/nsx034. URL: https://academic.oup.com/scan/article/12/7/1063/3067514.

[24] Prashant Doshi and Piotr J. Gmytrasiewicz. "Approximating state estimation in multiagent settings using particle filters". In: *Proceedings of the International Conference on Autonomous Agents* (2005). ISBN: 1595930949, pp. 463–470. DOI: 10.1145/1082473.1082522.

[25] P. Doshi et al. "Modeling Human Recursive Reasoning Using Empirically Informed Interactive Partially Observable Markov Decision Processes". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42.6 (2012), pp. 1529–1542.

[26] Dale O. Stahl and Paul W. Wilson. "On Players Models of Other Players: Theory and Experimental Evidence". In: *Games and Economic Behavior* 10.1 (July 1995), pp. 218–254. ISSN: 08998256. DOI: 10.1006/game.1995.1031. URL: https://linkinghub.elsevier.com/retrieve/pii/S0899825685710317.

[27] Miguel A Costa-Gomes and Vincent P Crawford. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study". In: *American Economic Review* 96.5 (Nov. 1, 2006), pp. 1737–1768. ISSN: 0002-8282. DOI: 10.1257/aer.96.5.1737. URL: https://pubs.aeaweb.org/doi/10.1257/aer.96.5.1737.

[28] P. Doshi and D. Pérez. "Generalized Point Based Value Iteration for Interactive POMDPs". In: *AAAI*. 2008.

[29] Prashant Doshi and et al. *A Particle Filtering Algorithm for Interactive POMDPs*. 2004.

[30] Michael L. Spezio. "Brain and Machine: Minding the Transhuman Future". In: *Dialog* 44.4 (Dec. 2005), pp. 375–380. ISSN: 0012-2033, 1540-6385. DOI: 10.1111/j.0012-2033.2005.00281.x. URL: http://doi.wiley.com/10.1111/j.0012-2033.2005.00281.x.

[31] Michael L. Spezio. "The neuroscience of emotion and reasoning in social contexts: Implications for moral theology: The Neuroscience of Emotion and Reasoning". In: *Modern Theology* 27.2 (Apr. 2011), pp. 339–356. ISSN: 02667177. DOI: 10.1111/j.1468-0025.2010.01680.x. URL: http://doi.wiley.com/10.1111/j.1468-0025.2010.01680.x.

[32] Charles A. Holt and Susan K. Laury. "Risk aversion and incentive effects: New data without order effects". In: *American Economic Review* 95.3 (2005), pp. 902–904. ISSN: 00028282. DOI: 10.1257/0002828054201459.

[33] Glenn W. Harrison et al. "Risk aversion and incentive effects: Comment". In: *American Economic Review* 95.3 (2005), pp. 897–901. ISSN: 00028282. DOI: 10.1257/0002828054201378.

[34] Daniel Kahneman and Amos Tversky. "Prospect theory: an analysis of decision under risk". In: *Econometrica* 47 (1979), pp. 263–291.

[35] Rui Alexandre Alves, São Luís Castro, and Thierry Olive. "Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill". In: *International Journal of Psychology* 43.6 (Dec. 2008), pp. 969–979. ISSN: 0020-7594, 1464-066X. DOI: 10.1080/00207590701398951. URL: http://doi.wiley.com/10.1080/00207590701398951.