Modeling cooperative and competitive decision-making in the Tiger Task

Saurabh Kumar
Institute of Systems Neuroscience
University Medical Center Hamburg, Germany
s.kumar@uke.de

Tessa Rusch
Institute of Systems Neuroscience
University Medical Center Hamburg, Germany
t.rusch@uke.de

Prashant Doshi
Department of Computer Science
University of Georgia, GA, USA
pdoshi@cs.uga.edu

Michael Spezio
Psychology & Neuroscience
Scripps College, CA, USA;
Institute of Systems Neuroscience
University Medical Center
Hamburg, Germany
mspezio@scrippscollege.edu

Jan Gläscher
Institute of Systems Neuroscience
University Medical Center
Hamburg, Germany
glaescher@uke.de

Abstract

The mathematical models underlying reinforcement learning help us understand how agents navigate the world and maximize future reward. Partially observable Markov Decision Processes (POMDPs) – an extension of classic RL – allow for action planning in uncertain environments. In this study we set out to investigate human decision-making under these circumstances in the context of cooperation and competition using the iconic Tiger Task (TT) in single-player and cooperative and competitive multi-player versions. The task mimics the setting of a game show, in which the participant has to choose between two doors hiding either a tiger (-100 points) or a treasure (+10 points) or taking a probabilistic hint about the tiger location (-1 point). In addition to the probabilistic location hints, the multi-player TT also includes probabilistic information about the other player's actions. POMDPs have been successfully used in simulations of the single-player TT. A critical feature are the beliefs (probability distributions) about current position in the state space. However, here we leverage *interactive POMDPs* (I-POMDPs) for the modeling choice data from the cooperative and competitive multi-player TT. I-POMDPs construct a model of the other player's beliefs, which are incorporated into the own valuation process. We demonstrate using hierarchical logistic regression modeling that the cooperative context elicits better choices and more accurate predictions of the other player's actions. Furthermore, we show that participants generate Bayesian beliefs to guide their actions. Critically, including the social information in the belief updating improves model performance underlining that participants use this information in their belief computations. In the next step we will use I-POMDPs that explicitly model other players as an intentional agents to investigate the generation of mental models and Theory of Mind in cooperative and competitive decision-making in humans.

Keywords: Theory of Mind, Tiger-task, Cooperation, Competition, Bayesian modeling, I-POMDP

Acknowledgements

J. G. was supported by the Bernstein Award for Computational Neuroscience (BMBF 01GQ1006) and J.G. and M.S. were supported by a Collaborative Research in Computational Neuroscience (CRCNS) grant (BMBF 01GQ1603; NSF 1608278). T.R. was supported by a PhD scholarship from the German National Merit Foundation.

Extended Abstract

1 Introduction

Reinforcement learning (RL) has its roots in artificial intelligence, control theory, operation research and has proven to be a powerful framework for cognitive neuroscience decision-making under uncertainty. *Markov decision processes* (MDPs) - the mathematical model underlying RL - help robots to pursue the goal of maximized total future reward by guiding their decisions when the state space is fully known. The real world, however, is imperfect with noisy observations and unexpected environment changes, where the current state of the world is often uncertain. The *partially observable Markov decision processes* (POMDPs) extend MDPs for situations of state uncertainty by proposing a belief distribution over possible states and using Bayesian belief updating for estimating this belief distribution in each moment (Kaelbling, Littman and Cassandra 1998).

The iconic Tiger Task played a crucial role in developing this computational framework by providing a test bed for simulating decision-making of a single agent in an uncertain world. The task mimics the setting of a game-show, in which the agent is presented with two doors, one of which hides a tiger (incurring a large loss) and the other one hides a pot of gold (incurring small win). The POMDP framework has been subsequently extended for multi-agent settings resulting in *interactive partially observable Markov decision process* (I-POMDP) (Gmytrasiewicz und Doshi 2005), in which two or more agent interact in an uncertain world. A crucial element of this framework is that agent build models of the other players and use them to predict others' choices and make better decisions themselves. The Tiger Task was again used in initial simulations of agents an their mental contents in this interactive setting.

Given the crucial role of the Tiger Task in formulating POMDPs and I-POMDPs it is surprising that little empirical data exist on this this task. Here, we set out to fill this gap by collecting choice data from human participants engaging in the single- and multi-agent Tiger Task, the latter being the focus of this paper. Furthermore, following Doshi (Doshi 2005) we devised a cooperative and a competitive version of the multi-agent Tiger Task and exposed two groups of subjects to them. In a series of model-free and model-based analyses of behavioural choice patterns we demonstrate a cooperative context elicits better choices and accurate predictions of the other player's actions and that subjects generate Bayesian beliefs to guide their actions. Critically, including the social information from the other player in the belief updating improves model performance, which underlines that participants pay attention to the other player and use it in formulating beliefs about the state of the world.

2 Task and hypothesis

The goal of the Tiger Task is to maximize the reward by opening the door hiding the gold (+10 point) and to avoid opening the door with the tiger (-100 points). In each step there are 3 actions available to the participant: open left door (OL), open right door (OR), or listen (L), which results in a probabilistic hint about the location of the tiger (growl left (GL), or growl right (GR)), but also costs 1 point. Thus, participant can accumulate evidence about the tiger location through repeated L actions. After each open action the position of the tiger is reset randomly to one of the two doors (tiger left (TL) or tiger right (TR)).

In the multi-player version the participants receive an additional probabilistic hint about the actions of the other player: creak left, or creak right (indicating that the other player might have opened one of the doors), or silence (S) indicating that the other player probably listened. Creaks suggest that the location of the tiger might have reset and that currently accumulated beliefs about the tiger location are void. Opening the door reveals the correct location of the tiger and the participant get the associated reward with additional knowledge of the tiger reset. In our implementation of the Tiger Task participant were also asked to predict the other player's actions at each step before choosing their own action (see Figure 1A for task sequence).

The competitive and cooperative versions differ in the structure of the payoff matrix: while the cooperative version incentivizes concurrent open actions by both players (see Figure 1C bold marking), the competitive version provides the maximum reward, if the correct door hiding the gold is opened, while the other player opens the wrong door hiding the tiger (see Figure 1B bold marking). Comparing the two versions,

we expected that participants will take more hints to come to a consensus in cooperative context to avoid confusing the other player and generate a more predictable behavior. We also expected more identical actions and more accurate predictions of the other player's actions during cooperation.

3 Results

We invited 58 participants (30 cooperate, 28 compete) to play the multi-player version of the game. In the model-free analysis we observed that the participants in the cooperative context took more hints than in the competitive context. In addition, prediction accuracy was higher during cooperation. These outcomes were both in line with our expectations. Participants in the competitive version exhibited fewer identical actions when compared to cooperation (Figure 2A-C).

Participants in the Tiger Task form beliefs about the states of the game (TL or TR) based on the probabilistic hints (GL or GR) and – in the multi-agent Tiger Task – the information from the other player (CR or CL). Because there are 3 distinct actions (OL, OR, L) available, we decided to model the action $a(t)$ at each step t as an ordered logistic regression model: $a(t) = ß_0 + ß_1 * b(t)$, where $b(t)$ is the belief about the location of the tiger.

The Tiger Task has only 2 states (TL and TR), which implies a unidimensional belief distribution with both states at the end of the range of possible beliefs. This belief distribution is updated on every step with the observations following the current action. We compared two version of belief updating: a simple "beta-belief" model, which uses the mode of a beta distribution as the point estimate of the belief and is updated by adjusting the parameters of the beta distribution with the observations (the probabilistic hints following L actions). The second model is a Bayesian belief updating model with take the previous belief as the prior and calculates the likelihood based on the observation and transition function. We also tested two versions of the Bayesian updating model (*Eq 1*) without and E*q 2*) with the inclusion of the social information (also see Figure 3A-B as an example):

$$b(t) = \frac{p(gc)*b(t-1)}{p(gc)*b(t-1)+(1-p(gc))*(1-b(t-1))} \quad (1)$$

Where, p(gc) is the probability of the hint being correct and b(t-1) is the previous belief about the tiger location.

$$b(t) = \frac{p(cc)*p(oo)}{p(cc)*p(oo)+(1-p(cc))*(1-p(oo))} * p(reset) + 1 - \frac{p(cc)*p(oo)}{p(cc)*p(oo)+(1-p(cc))*(1-p(oo))} * b(t-1) \quad (2)$$

Where, p(cc) is the probability of the hint about the partners' action being correct and p(oo) is the probability of the partner opening the door, while reset is the probability of the tiger being placed after a door is opened (0.5 for a random placement).

Models were estimated using the Stan software package that implements and hierarchical Bayesian workflow. Formal model comparison using LOOIC (Leave-one-out information criterion) revealed that the Bayesian belief update model resulted in a better fit than the beta-belief model (LOOIC (Bayesian belief) = 5107.75, LOOIC (Beta belief) = 8530.70). In control analysis, we expanded the set of predictors in the ordered-logistic model with additional task variables like the number of hints taken, previous outcome and an interaction between them (Model 2-5), but found the simpler model with just the belief as a predictor (Model 1) outperforms these more comprehensive predictor sets (Figure 4A-B). Furthermore, we compared the Bayesian belief update without the social information (Eq 1) to the update with the social information added (Eq 2) and concluded that the social information adds a significant improvement in the model prediction (see the scales of LOOIC values in Figure 4A and 4B).

4 Outlook

We used an ordered logistic discrete choice model with Bayesian belief updating and demonstrated that including the social information is providing a much better model fit to the data. This suggests that participants in the multi-agent Tiger Task do incorporate the information from the other player into their valuation process. However, our Bayesian belief model falls short of an important feature that is likely shaping strategic social decisions: it treats the information from the other players as just another piece of information from the environment and not as an intentional agent that processes the information in a similar way.

I-POMDPs are a computational framework that explicitly computes the beliefs of the other player as an intentional agent as part of the model of the first player. Thus, it is an ideal framework for modeling Theory of Mind of another player in a quantitative way (his goals, intentions, and beliefs). Following our Bayesian belief model, we will also model the Tiger Task within the I-POMDP framework and compare belief computations of the other player in the competitive and cooperative version of the task.

5 References

Doshi, Prashant. „Optimal sequential planning in partially observable multiagent settings." Ph.D. dissertation, University of Illinois at Chicago, 2005.

Gmytrasiewicz, Piotr J., und Prashant Doshi. „A framework for sequential planning in multi-agent settings." *Journal of Artificial Intelligence Research*, 2005.

Kaelbling, Leslie Pack, Michael L. Littman, und Anthony R. Cassandra. „Planning and acting in partially observable stochastic domains." *Artificial Intelligence*, 1998.

6 Figures

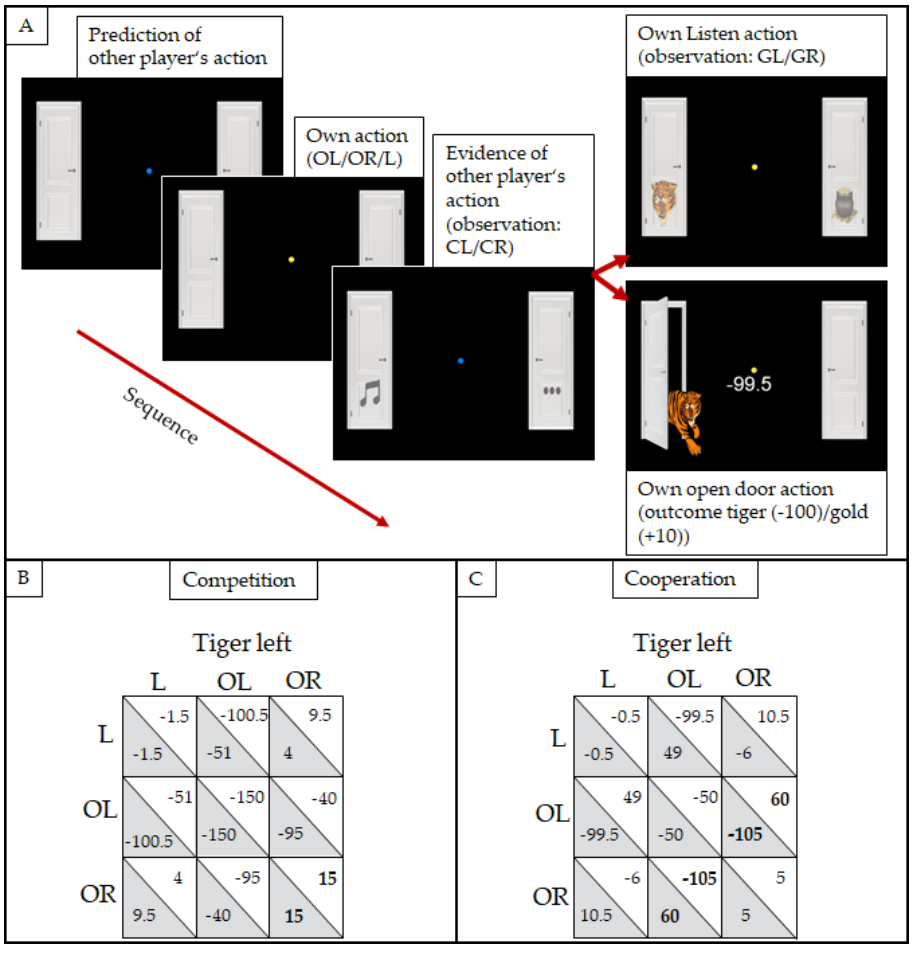

Figure 1: (A) An example of sequence for the multi-player version of the task. The player predicts the action of the other player (indicated by the blue fixation dot) followed by the players own action choice (indicted by the yellow fixation dot). This is followed by the probabilistic evidence about the other player's action (CL/CR). The next screen is either the probabilistic hint about the tiger location (GL/GR if L was chosen) or the door is opened (for OL or OR actions) revealing the tiger location. (B) The joint payout matrix in the competitive context is shown for the tiger being on the left side. The bold numbers show the best and worst choice indicating that the best own outcome is achieved if the correct door (with the gold) is opened, while the other player open the wrong door (with the tiger). (C) The joint payout matrix in the cooperative context is shown when the tiger is on

the left side. The bold numbers showing the best choice indicating that the maximum payoff is achieved, when both players open the correct door at the same time.

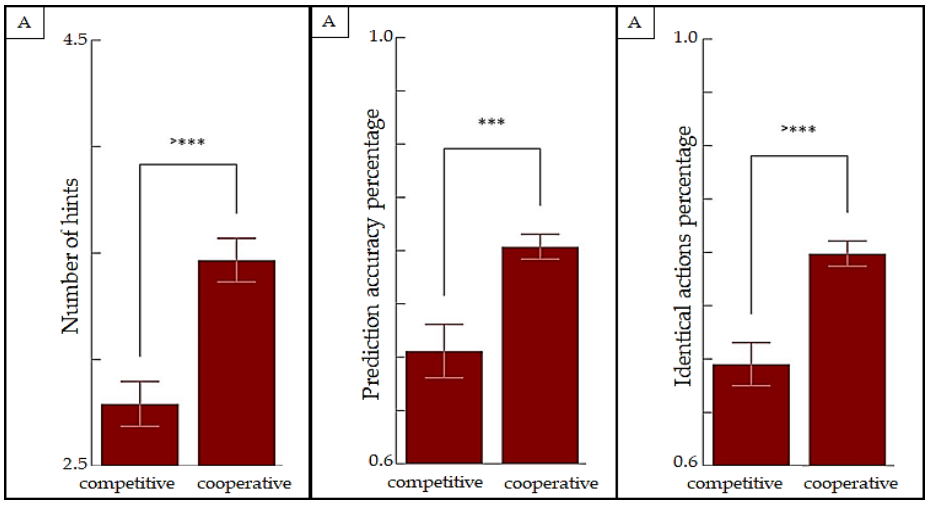

Figure 2: (A) In the multi-player version of the task the participants significantly took more hints in the cooperative context when compared to the competitive context. Participants also had significantly higher prediction accuracy and identical actions (showing coordination) in the cooperative context compared to the competitive one in (B) and (C) respectively.

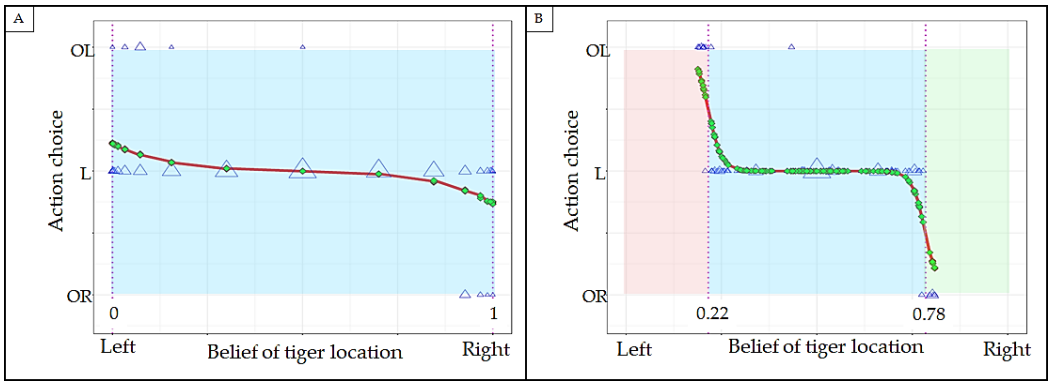

Figure 3: (A) An example model behavior of the multi-player version of the tiger task without the social information (see Eq 1) is shown here. The bold red line is the model prediction while the blue triangles are the actual participant action choices given their computed beliefs. Green dots, which always lie on the red model curve show the model predictions of the data (blue triangles). The light-blue area shows the belief region where the ordered logistic model predicts the listen action. In the red and green areas the ordered logistic model predicts Open Left and Open Right action respectively. The absence of these areas in this model suggests that the model without the social information fails to predict the observed open left/right actions. (B) This model behavior shows the prediction made with the social information (Eq 2). This model predicts most of the OL actions (red area) and OR actions (green area) correctly demonstrating the importance of the social information (CR/CL) for correctly predicting the observed data.

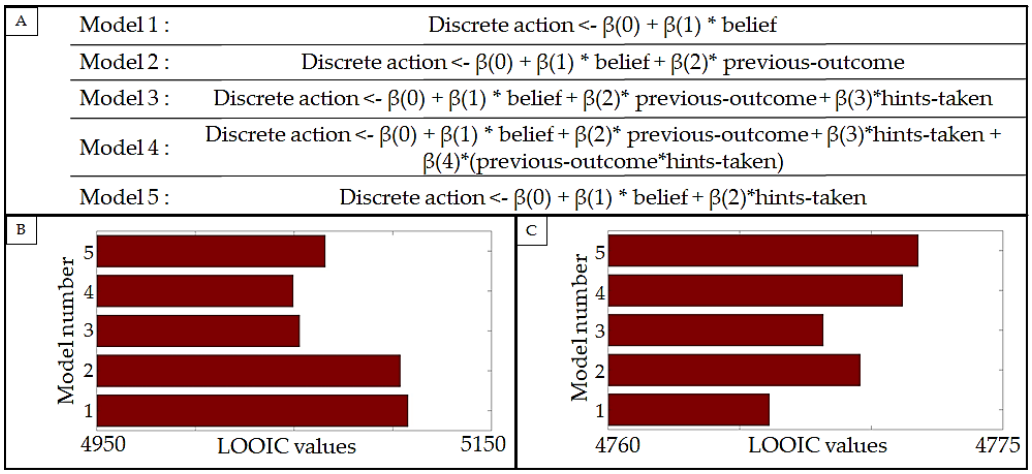

Figure 4: (A) Different models compared of the multi-player version of the task. All the models in (B) without the social information perform worse compared with the LOOIC values of the models with the social information added in (C) (see different scales in (B) and (C)). The simplest model with just the belief update (model number 1) in (C) performed better when compared to extensions of number of hints taken, previous outcome and an interaction of them.