A Neuro-Computational Characterization of Theory of Mind Processes during Cooperative Interaction

Tessa Rusch (t.rusch@uke.de)
Inst. for Systems Neuroscience, University Hamburg
Martinistr. 52, 20246 Hamburg, Germany

Prashant Doshi (pdoshi@cd.uga.edu)
Dept. of Computer Science, University of Georgia
Boyd GSRC, Athens, GA 30602, USA

Martin Hebart (martin.hebart@nih.gov)
National Institute of Mental Health
10 Center Drive, Bethesda, MD 20814, USA

Saurabh Kumar (s.kumar@uke.de)
Inst. for Systems Neuroscience, University Hamburg
Martinistr. 52, 20246 Hamburg, Germany

Michael Spezio (mspezio@scrippscollege.edu)
Psychology & Neuroscience, Scripps College
1030 Columbia Ave., Claremont, CA 91711, USA

Jan Gläscher (glaescher@uke.de)
Inst. for Systems Neuroscience, University Hamburg
Martinistr. 52, 20246 Hamburg, Germany

Abstract

Humans are distinctly skilled at cooperation. To successfully engage with others they apply Theory of Mind (ToM). Here, we investigate neuro-computational mechanisms underlying ToM during real-time dyadic coordination in a probabilistic social decision game. To effectively coordinate participants have to represent the surrounding they interacted in and simultaneously simulate their partner's representation of the world. These cognitive computations are formalized with a decision framework that combines decision-making under uncertainty with intentional models of other agents. Using model-based EEG analyses, we identify oscillatory signals related to errors experienced by players when own expectations towards the surroundings are violated and simulations of errors experienced by the partner when the partner's predictions fail. Consistent with previous studies, we find positive correlations between power in frontal delta and theta oscillations and experienced errors. Most strikingly, these signals are also found in relation to simulations of the partner's error, at times when participants themselves experience no prediction error themselves. These findings unveil the neural signature of a crucial computational component of the mental model of a partner and demonstrate that the brain recruits similar mechanisms for simulation the decisions of others as for computing one's own decision.

Keywords

Theory of Mind; cooperative decision making; EEG, cognitive modelling; I-POMDP

Background and aim

Humans are experts in cooperating with others. Cooperation is the capacity to act in accordance with the percepts, goals, and beliefs of others to facilitate own and other's gain equitably. Cognitively, it requires Theory of Mind (ToM), i.e. the ability to estimate and represent others' mental states and predict rational behavior based on these mental states. To successfully cooperate, humans have to combine the predictions of a partner's behavior with their knowledge of the world and act according to the combined requirements of the interactive situation. Here, we set out to investigate the neuro-computational mechanisms allowing humans to cooperate by formalizing the models humans build of others' mental states and the world and identifying neural signals related to updating of these models.

To examine the processes underlying cooperative behavior in a truly interactive setting we developed a decision-making task that requires cooperative choices on probabilistic and occasionally changing options. Action planning is complicated by the facts that participants receive only noisy observations of the underlying task structure, which on top can change unpredictably. A cooperative reward structure incentivizes coordination. Additionally, one player has more knowledge about the task situation than the other. The asymmetry of knowledge between two agents resembles the situation in the classic false-belief-task (Wimmer & Perner, 1983), in which the (all-knowing) participant has to realize that the girl in the story has a false belief about the environment and therefore makes an incorrect choice. In our setup, the less informed player has a false belief and the informed player a correct belief about the world. As the joint reward is maximal when choices are coordinated, divergent knowledge prompts participants to observe and learn about the surroundings *and* model and track the partner's belief about the world. To act optimally, they need to combine their world knowledge and their model of the partner's mental state in a single valuation process.

To gain access to the private cognitive operations that allow humans to coordinate in a complex setting, we model behavioral data in the context of the I-POMDP framework (Gmytrasiewicz & Doshi, 2005). I-PODMPs extend single-agent action planning in an uncertain environment to the interactive domain by including intentional models of other agents that themselves engage in action planning. These models of others may themselves include models of the original agent allowing the capture of recursive reasoning processes humans can engage in during strategic interaction.

Building on the cognitive model we examine participants' neural signals recorded with EEG in a model-based approach. We aim at identifying neural signals related to the updating of players' models of the world as well as updating processes related to simulating the partner's model of the world.

Methods and results

Task Details

The task used here extends the concept of the classic false belief task to the interactive domain. We therefore refer to it as the "Interactive False Belief Task" (IFBT). In the IFBT two players choose between two options ("left" or "right") for probabilistic rewards. One option has a high probability for a high reward (10), the other a high probability for a low reward (5). Using trial and error participants find out whether the high reward is on the left or on the right. When both partners obtain the
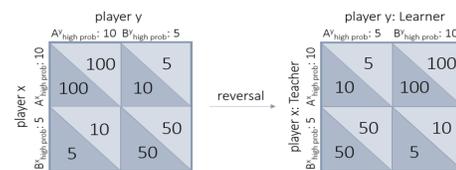

Figure 1| Pre- and post-reversal reward matrices

same individual outcome, they are rewarded by a ten-fold increase of their individual outcomes. If individual outcomes differ, they receive the nominal individual outcomes. Their own reward distribution and the partner's action are unknown to the players, but have to be inferred form the received outcome. The partner's reward distribution is openly presented to the players at the beginning of each trial. Prior to their own choice, participants have to predict the partner's action. In the displayed reward matrices (Figure 1), the initial setting is shown on the left. Both players need to choose option "A" to receive the individual high outcome. Thereby, the probability of receiving the maximum reward of 100/100 is highest. However, due to the probabilistic choice-outcome relation, all other outcomes are also possible. After a few trials, one player's (here: Player y's) reward contingencies are reversed, i.e. this player's high option moves from left to right or vice versa, while the partner's reward contingencies remain the same. This player remains uninformed about the change and is therefore referred to as the "Learner". As Learners are ignorant to reversals, they hold a false belief about the reward structure of the task. The partner is informed about the contingency reversal, hence we call this player "Teacher". For ease of reference we will refer to the Teacher as "she" and the Learner as "he". This is unrelated to the participants' gender, as we tested an equal number of male and female participants and all participants played both roles in exclusively same-gender dyads (total N = 50, 25 female). Taking the Learner's false belief into account the Teacher has to choose the less valuable option "B" at reversal. The most likely ensuing reward of 50/50 signals the Learner that his reward contingencies have reversed. After a period of stable coordination, reversals repeat. Throughout the game, reversals are unpredictable and players are randomly assorted to the roles of Teacher and Learner.

Teacher and Learner predictions and choices

The Learner's main task in the IFBT is to detect and react to changes in the reward-contingencies. The Teacher is fully informed about the change. Her goal is to "communicate" these reversals through her choices. She has to react to the Learner's decisions at the reversal and to his choice adaptation after the reversal.
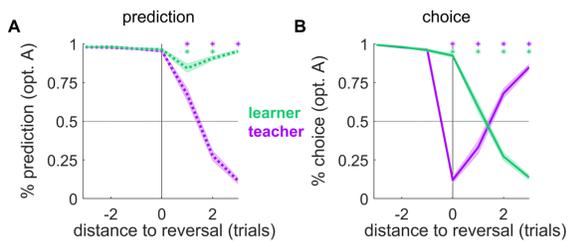
Figure 2l Teacher and Learner predictions and choices

Players' predictions about the partner's choices are shown in Figure 2A, players' own choices in Figure2B: The Learner detects reversals and gradually shifts his choices after a reversal. During the reversal, the Teacher correctly predicts that the Learner stays with his previous choice, but switches her own choice to "B", the Teacher's individually less rewarding choice option, which is, however, still the best option given the reversal in the Learner's reward contingencies. In post reversal trials, the Teacher accurately predicts the partner's choice curve (purple prediction curve in (A) and green choice curve in (B)) and matches the Learner's choice switching by returning to her pre-reversal choice at the same rate (purple choice curve in (B)). These results strongly suggest, that participants actively engage in mentalizing to solve the task.

Modelling interactive decisions

In the IFBT outcomes are probabilistically associated with the participants' choice options. The goal of the task is to maximize the joint outcome. To achieve this goal participants have to generate beliefs about which option is currently the best. Based on their actions and observations of the resulting joint outcome they can update this belief distribution. After a reversal, the Learner does not know that the reward situation has changed. His belief is therefore false. The Teacher is aware of the change. Furthermore, she knows that the Learner does not know about the reversal. Thereby, she can infer the Learner's false belief, correctly predict the Learner's (wrong) action, and accommodate for it by switching her own choice. Previous studies examining ToM in interactive tasks did not include uncertainty about the surrounding (e.g. Hill et al., 2017; Yoshida, Dolan, & Friston, 2008). In these studies, representing another persons' beliefs is unnecessary, as in a fully and perfectly observable world, others' beliefs should be identical to one's own belief. In the current study, however, participants interacted in a highly uncertain environment. Therefore, we need to address the attribution of beliefs to others, a core component of ToM.

Single agent action planning under uncertainty is well captured by partially observable Markov Decision Processes (POMDPs) (Kaelbling, Littman, & Cassandra, 1998). The innovative element of POMDPs is that the agent maintains a belief about the world. Beliefs are represented by probability distributions over all possible discrete states of a world. At each time step the agent's belief is updated with a Bayesian learning rule. In the context of the IFBT, states are specified by the location of the high reward option (possible states are "High Left (HL)" and "High Right (HR)"). Here, we extend the problem to the multi-agent domain.

To capture humans mentalizing during interaction in an uncertain environment, we apply Interactive POMDPs (I-POMDPs) (Gmytrasiewicz & Doshi, 2005). In contrast to single agent POMDPs, I-POMDPs contain an agent's belief about the states of the world and a belief about the mental states of the other agent, which is the other agent's belief about the states of the world. For the IFBT this means that agents form a belief about the location of their own and their partner's high option, and about the partner's belief about the distribution of rewards. As in the single agent model, beliefs are updated at each time step. In the multi-agent framework, the belief about the other's mental state is updated by simulating the partner's learning process. These core features of the framework make it an ideal candidate for modeling participants' behavior in the IFBT and access the underlying critical belief computations.

We fitted parametrized I-POMDP models to behavioral data from the IFBT and found that I-POMDPs predict the Teacher's and the Learner's actions with high accuracy (compare model predictions (dotes lines) and participants average behavior (solid black lines) in Figure 3). From the fitted I-POMDPs players' beliefs about their respective reward contingencies (HL/HR) and their belief about the partner's belief are computed. Players own beliefs are represented by the solid colored lines in Figure 3C and (D), their representations of the partner's belief in (A)
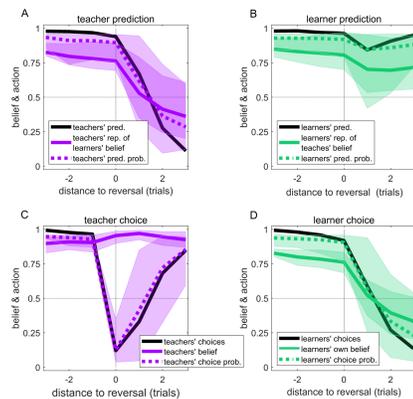

Figure 3l Model predictions and participants' actions

and (B). The Teacher's individual reward distribution remains constant throughout the peri-reversal period and extracted beliefs show that the Teacher correctly represents the reward structure (C). Before a reversal, also the Learner correctly represents the state of the task (D). However, at reversal, the reward distribution changes but the Learner's belief stays constant, i.e. becomes false. In the post-reversal period, his belief starts to shift reflecting the gradual adaptation to the change in reward contingencies. The Teacher's belief about the Learner's belief matches the Learner's own beliefs almost perfectly (compare solid purple and green line in (A) and (D)). This shows that Teachers accurately represent their partner's mental state.

Model-based EEG analyses

From extracted beliefs we computed players' trial-wise own expectations towards outcomes as well as the Teacher's simulation of the Learner's expectations towards outcomes. Based on these expectations we deduced two different prediction errors (PEs): An experienced PE, capturing the surprise experienced by players in response to observed outcomes, and a simulated PE that represents the Teacher's simulation of the Learner's surprise. At and after reversals, the Learner's belief about the task situation is incorrect, hence he experiences strong PEs. The Teacher on the other hand correctly represents the reward structure and therefore experiences small PEs when outcomes are presented. However, in addition to experienced PEs, the Teacher simulates the Learner's surprise about outcomes. In line with the Learner's experienced PE, the Teacher's simulates large PEs at and after reversal. Using single trial regression analyses, we related experienced and simulated PEs to power of oscillatory neural signals. In line with previous research

(Cohen & Cavanagh, 2011), the Learner's experienced PEs correlate positively with power in the delta and theta band (left half of Figure 4). In addition, we find that also the Teacher's simulated PE positively correlates with low frequency signals similar to the Learner's experienced PE response (right half of Figure 4). These findings suggest that simulating a partner's dynamic mental state during coordinated decision making is instantiated by similar neural mechanisms as engaging in them oneself.

Discussion

Successful cooperation requires coordination of joint actions. Here we provide behavioral, computational and neural evidence that humans represent their partners as rational intentional agents and dynamically model their mental states. Using the I-POMDP framework we can formalize and quantitatively estimate these Theory of Mind processes. Our modeling findings suggests that humans incorporated mental models of their partners into their own model of the world and use it to guide coherent decision making. Further, we show that modelling a partner's decision process recruits similar neural mechanism as own action planning and decision making.

References

Cohen, M. X., & Cavanagh, J. F. (2011). Single-Trial Regression Elucidates the Role of Prefrontal Theta Oscillations in Response Conflict. *Frontiers in Psychology*, *2*(FEB), 1–12.

Gmytrasiewicz, P. J., & Doshi, P. (2005). A Framework for Sequential Planning in Multi-Agent Settings. *Journal of Artificial Intelligence Research*, *24*, 49–79.

Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O'Doherty, J. P., & Ruff, C. C. (2017). A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience*, *20*(8), 1142–1149.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game Theory of Mind. *PLoS Computational Biology*, *4*(12).
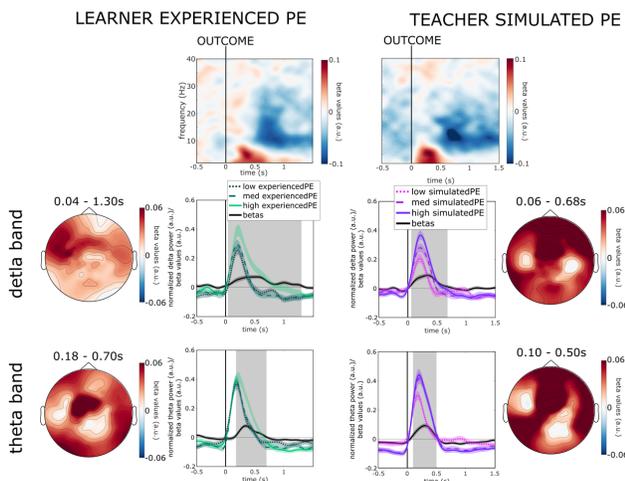
Figure 4l Experienced and simulated PE regression results