Reasoning about self and others

Meijering, Ben

Chapter 3

Reasoning about self versus others: Changing perspective is hard

Abstract

To understand others, we need to infer their mental states, such as beliefs, desires, and intentions. Some developmental studies have suggested that general reasoning ability plays a crucial role in inference of mental states. In contrast, we show that the most important factor for successful inference of mental states is the ability to change perspective. In our experiment, participants either had to make a decision, or predict how another person would make the exact same decision. Crucially, the required steps to solve both problems were the same. Nevertheless, participants made more mistakes and required more time while making predictions instead of decisions. This finding implies that perspective taking, while making predictions, employs computational processes that are unique to the mental aspects of a problem.

This chapter was submitted to a journal where it is currently under review.

We are living in a social world in which many of our daily activities involve interactions with others. These interactions can take many forms: negotiating a higher salary, gossiping with a friendly neighbor, bluffing in a card game, or having a conversation via Facebook or Twitter. Irrespective of form, a social interaction requires the ability to infer another's knowledge, beliefs, desires, and intentions. For example, if we are bluffing in a game of poker we are reasoning about what the other player may think our intentions are if we raise the bet. Reasoning about beliefs and other mental states requires a so-called Theory of Mind (ToM), which develops around the age of 4 (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). That is relatively late considering that infants as young as 7 months already seem to be susceptible to the beliefs of others (Kovács, Téglás, & Endress, 2010; also see O'Neill, 1996; Onishi & Baillargeon, 2005). Furthermore, the process of putting oneself in another's shoes seems to be almost effortless and automatic (Cohen & German, 2010; Kovács et al., 2010; Ramsey, Hansen, Apperly, & Samson, 2013; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). Given these findings, it is surprising that it takes so long before children learn to infer the beliefs of others and that adults still frequently fail to use ToM (Apperly et al., 2010; Keysar, Lin, & Barr, 2003; Lin, Keysar, & Epley, 2010).

In recent years, some developmental studies have investigated whether preschoolers' difficulties with ToM are specific to the domain of mental states (Leekam, Perner, Healey, & Sewell, 2006; Perner & Leekam, 2008; Sabbagh, Xu, Carlson, Moses, & Kang, 2006). They set out to investigate whether problems in ToM tasks arise because preschoolers have to put themselves into another's shoes, or because the underlying logical problems in these tasks require cognitive functions that have not yet matured. The preschoolers were presented with two tasks. In one, the false-belief task, the preschoolers were presented with a story in which a person, call her Anne, stored a toy in a box before leaving the room. While Anne was away, a second person, Bob, moved the toy to another location without Anne being aware of it. After Anne returned, the preschoolers were asked where she would look for the toy. To answer correctly they had to reason about Anne's false belief that the toy was still in the box. Performance in this task was compared with performance on the so-called false-sign task, which has the same logical structure but does not involve mental states. A toy is stored in one location, and a sign is pointing at that location. Next, the toy is moved to another location, but the sign is still pointing at the original location. After the story had been told, the preschoolers were asked to indicate where the toy should be according to the sign. In this task, children did not have to reason about mental states (i.e., a false belief), but they did have to refrain from indicating the actual location of the toy, similar to the false-belief task. Because performance in this task correlated with performance in the false-belief task, it seemed that preschoolers' difficulties in understanding false beliefs were not solely confined to mental states (Leekam et al., 2006; Sabbagh et al., 2006). In fact, performance on both tasks correlated with measures of general reasoning ability (Sabbagh et al., 2006), which suggests that inference of mental states suffered from limited capacity of general cognitive functions, not ToM proficiency (also see Bloom & German, 2000).

These findings may seem compelling, but why then do adults, whose general cognitive functions have matured, still frequently fail to apply ToM to interpret the behavior of others? Adults have sufficiently developed general cognitive functions to be able to comprehend many logical problems. Yet, it seems that they reflexively reason about their own beliefs when interpreting the behavior of others (Apperly et al., 2010; Keysar et al., 2003; Lin et al., 2010).

We therefore argue, in contrast to the developmental studies, that people *do* find it difficult to put themselves in another's shoes. The preschoolers in the developmental studies may have had such difficulties too, but they did not have sufficiently matured general reasoning ability to understand the logical structure of false beliefs and false signs, to begin with. In contrast, adults do have mature general reasoning ability and therefore are the appropriate population to investigate the role of perspective taking during inference of mental states: If adults understand the logical structure of a given ToM task but find it difficult to put themselves in another person's shoes, unsuccessful application of ToM would reflect a deficiency of a specialized function, not a deficiency of general reasoning ability.

To test specifically whether adults find it difficult to put themselves in another's shoes, we devised a task in which they either had to make a decision themselves, or had to predict how another would make the very same decision (Appendix A). The task consisted of two-player games in which the participant and a computer-simulated player alternately made decisions. There were two conditions and the only difference between the conditions was the required level of perspective taking; all other task aspects were the same. Importantly, this was a direct and specific test of perspective taking, because participants were not asked to reason about distinct types of representations, which was the case in the false-belief and false-sign tasks. Crucially, the steps to 'solve' the games were the same in both conditions, irrespective of the required level of perspective taking. The only difference for a participant was the instruction, or prompt, given at the start of each game. In one condition the prompt was "Decide" what

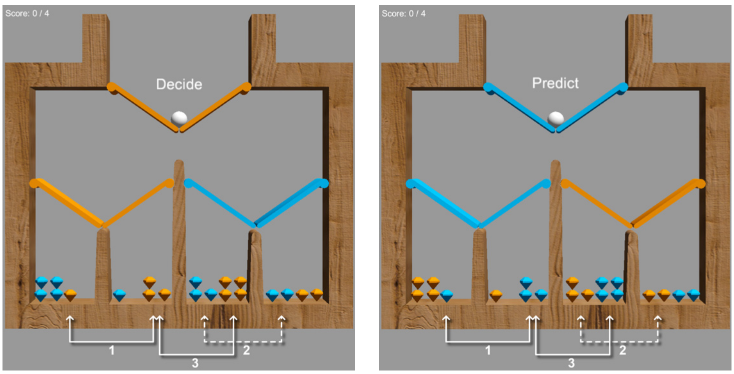

Figure 3.1: Screenshots of isomorphic Decision (left panel) and Prediction (right panel) games. The arrows were added for illustrative purposes and show that the games require the same comparisons to obtain the best possible outcome. In these particular games, the participant is assigned the target color orange and has to obtain as many orange diamonds as possible. The game in the left panel prompts the participant to make a decision: "Decide". To make a decision, the participant needs to switch perspective once, as his outcome depends on the decision of the other player, who decides at the bottom-right trapdoors and whose goal is to obtain as many blue diamonds as possible. The game in the right panel prompts the participant to predict the other player's decision: "Predict". The participant needs to switch perspectives twice: The first time to reason about the other player's intention at the topmost trapdoors, and a second time to switch back to his own perspective, because the other player's decision depends on the participants decision at the bottom-right trapdoors.

you would do in this game, and in the other condition it was "Predict" what the other player would do in this game (see Figure 3.1). Again, the steps to arrive at a decision or a prediction were the same, but it mattered whether participants were asked to execute these steps from their own perspective or from the other's, as we will show.

At the start of the experiment, each player was assigned their own target color (orange or blue) and the goal was to obtain as many target-colored diamonds as possible (Appendix A). In each game a white marble dropped into a contraption (see Figure 3.1) and both players could influence its path by opening the trapdoors depicted in their target color. Each player obtained the target-colored diamonds located in the bin into which the marble dropped. Both players had to reason about one another, because the other player's decisions affected their outcomes. In games such as the one in the left panel of Figure 3.1, for example, participants were prompted to make a decision at the topmost trapdoors. They could infer that their best possible decision is to open the right-side trapdoor, because the other player's intention is to open the left-side trapdoor, as his goal is to obtain as many blue diamonds as possible: He would obtain 3 blue diamonds, and the participant would obtain 4 orange diamonds. In these so-called Decision games participants had to switch perspective just once while they were reasoning about the other player. In so-called Prediction games, however, participants had to switch perspective twice (Figure 3.1; right panel): To predict the other player's decision at the topmost trapdoors, they had to switch from their own perspective to that of the other player. Furthermore, as the topmost decision was based on their own decision at the bottom-right trapdoors they had to switch back again from the other player's perspective to their own. In other words, participants had to reason about recursive mental states: "*the other player thinks* that *I intend* to open the left-side trapdoor, as *he knows* that *my goal* is to obtain the highest possible number of orange diamonds". Importantly, the Decision and Prediction games were structurally equivalent: The order of the decisions was the same in both types of games, as was the distribution of diamonds. Thus, participants had to make the same comparisons between the same distributions of diamonds, as can be seen in Figure 3.1, irrespective of the prompt given at the start. The only difference between the games was the perspective from which to make the comparisons.

In this study, we are comparing two hypotheses. Based on the earlier mentioned developmental studies, one would expect the response patterns to be the same in Decision and Prediction games, because these games require the same steps to 'solve' them (see Figure 3.1). As a consequence, Decision and Prediction games induce the same demands on general reasoning ability. The accuracy and response times are therefore not expected to differ, which makes this the null-hypothesis. The other hypothesis is that adults will perform differently in Decision and Prediction games, because these types of games differ with respect to the required level of perspective taking. A previous study has suggested that people do switch between perspectives each time they fixate a new set of trapdoors (Meijering, Van Rijn, Taatgen, & Verbrugge, 2012). Based on that study, we expect the accuracy and response times to differ, because the number of switches between perspectives differs between Decision and Prediction games.

As can be seen in Figure 3.2, both the accuracy (i.e., the proportion of correct responses) and the response times indicate that it is more difficult to reason about someone else's decision-making than making the same decisions oneself (Appendix B). In the Prediction games, in comparison to the Decision games, accuracy is significantly lower, $\chi^2(1) = 44.77$, p

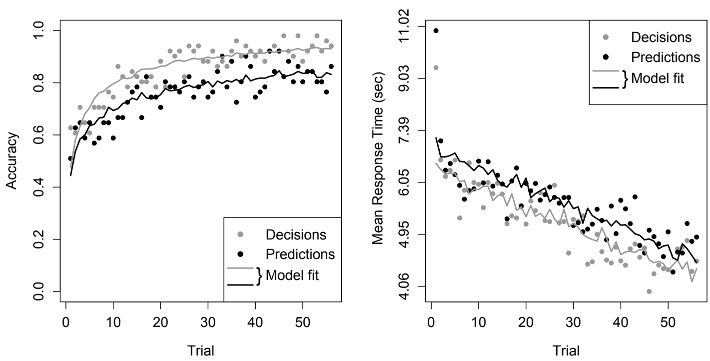

Figure 3.2: The left panel depicts the mean accuracy of decisions (grey) and predictions (black) across participants. The right panel depicts the actual response times on a logarithmic scale, also averaged across participants. The solid lines depict fits of linear mixed-effects regression models.

< 0.001, and the response times are significantly longer, $\chi^2(1) = 4.97$, $p < 0.05$. These findings imply that the reasoning processes were not symmetrical in the two types of games. We argue that the asymmetrical response patterns are due to differential demands on perspective taking, because the games are isomorphic otherwise. In prediction games participants had to switch perspectives twice, and as a consequence they made more mistakes and needed more time to produce a response. We accounted for the fact that Decision and Prediction games were presented in an intermixed fashion, which may have caused asymmetric reorientation costs while switching back and forth between Decision and Prediction games. However, this reorientation factor was not significant, nor was any interaction with it. Thus, the demand on perspective taking was the sole factor determining accuracy in Decision and Prediction games.

There is a 'smart' strategy to play these games, and that is to pretend to switch between colors, reasoning as if one is the 'orange' player in some games (e.g., Decision games) and the 'blue' player in the others (e.g., Prediction games). By using this strategy the participants would have reduced the required level of perspective taking of all games to level one. Strikingly, the participants did not use this strategy, as the accuracy and response times differ between Decision and Prediction games. Our findings therefore show that the participants were strongly committed to their own target color and reasoned from their own perspective when prompted to "Decide" and from the other player's perspective when prompted to "Predict". In fact, the difference in accuracy and the ratio of associated response times did not become any smaller during the experiment (Appendix B). This is remarkable as participants were presented with 112 games in total, which is ample opportunity to adopt a smart strategy. The fact that the differences did not become smaller is a strong indication that the participants committed to their own target color and engaged in perspective taking, which caused differential demands on mental state reasoning in the Decision and Prediction games.

There are several possible explanations for differential response patterns in Decision and Prediction games. For example, the process of modeling the other player's mental states may require cognitive functions that are not as well developed as the cognitive functions to model one's own mental states (Birch & Bloom, 2004). The process of modeling the other player's goals and intentions could therefore be prone to errors, yielding incorrect predictions. A related explanation is that people fall prey to the egocentricity bias and consequently interpret the behavior of the other player according to their own goals and intentions (Birch & Bloom, 2007; Keysar et al., 2003). Overcoming this bias may be difficult and cause high cognitive demands, as one has to inhibit one's own mental states (German & Hehman, 2006; Samson, Apperly, Kathirgamanathan, & Humphreys, 2005). Lastly, a representation of the other player's mental states could intrinsically be more complex (Gopnik & Wellman, 1992), because the mental states have to be labeled as belonging to that player. One's own mental states, in turn, do not have to be labeled. In any case, each of these possible factors is a consequence of perspective taking, as perspective taking alone distinguishes between Decision and Prediction games.

To conclude, the results show that reasoning about someone else's decision-making is more difficult than making the same decisions oneself, even if the conditions are equivalent. The steps to arrive at either a decision or a prediction were the same in this study, but apparently participants engaged in distinct processes. Critically, predictions required more switches between perspectives than did decisions, and as a consequence participants produced fewer optimal responses and required longer response times. This study shows that the bottleneck of mental state reasoning is perspective taking: Reasoning about self is not as difficult as reasoning about others.

Appendix A: Methods

Ethical statement

The study was approved by the ethical committee of the Psychology department of the University of Groningen.

Participants

In this study, 51 first-year psychology students (34 female) participated in exchange for course credit. Their mean age was 21, ranging from 18 to 34 years. None of the participants were excluded from the data analyses. All participants had normal or corrected-to-normal visual acuity. Written informed consent was obtained from all participants.

Stimuli

Each game had a unique distribution of payoffs (i.e., diamonds). Of all possible payoff distributions we excluded those that did not require inference of mental states. There were three exclusion criteria in total. First, a payoff distribution was excluded if the player deciding at the topmost trapdoors had her two highest payoffs on one side and her two lowest payoffs on the other side. In this case, the player would not have to consider the other player's decision. Second, a payoff distribution was excluded if the maximum payoff of the player deciding at the topmost trapdoors was behind her own two sequential trapdoors. In case of such a payoff distribution, she does not need to reason about the other player's goals and intentions. Third, a payoff distribution was excluded if both players had the same number of diamonds in each bin. Such a payoff distribution does not require inference of mental states, as there would not be any conflict between the two players' goals, beliefs, and intentions.

From the remaining 192 payoff distributions, 56 were randomly selected to be included as the set of Decision games. Another 56 items, also randomly selected from the 192 payoff distributions, comprised the set of Prediction games.

Procedure

The participants were seated in front of a 24" monitor on which the games were played. Before the task started, instructions were given on paper. The instructions explained: the goals of the participant and computer-simulated player; who decided where; and what response participants should give in case of the "Decide" and "Predict" prompts. After reading the instructions, participants could ask for clarification if they had any questions. The experimenter answered these questions, but was careful not to give any information on the strategy of the computer-simulated player.

The decision and predictions games were presented in random order, and the participants played these games from start to end, until the marble dropped into one of the bins. The games were fully animated, and at the end of each game there was feedback, which indicated the participant's score in that particular game. For example, "+3" if the marble dropped into a bin that contained three diamonds in the participant's target color. The total score was depicted in the top left corner of the screen. During the first twelve games, participants were provided with additional feedback after each game, which indicated whether the obtained score was the highest possible score they could have obtained in that particular game.

Appendix B: Results

The individual responses (decisions and predictions) were analyzed by means of *logistic* linear mixed-effect models, as the data contained at least two sources of random variation: The participants and payoff structures were both sampled from larger populations. We constructed a full factorial model that comprised fixed effects of ToM *order* (level 1 / level 2), *switching* between ToM-orders (switch / no-switch), and the covariate *trial*. Trial was log-transformed to account for the non-linear increase in the proportion of correct decisions and predictions (see Figure B.1). The values of the *Akaike information criterion* (AIC) and the *Bayesian information criterion* (BIC) of the full factorial model were compared with those of simplified models from which interaction and main effects were removed. The AIC and BIC values indicate whether a better fit of a more complex model is justifiable given its extra parameters.

Accuracy

The model with the most favorable (i.e., smallest) AIC and BIC values contained main effects of *order* and *log-trial*, and an interaction between the two. The absence of main effects and interaction effects on accuracy of switching between Decision and Prediction games, suggests that reorientation costs, if any, were negligible and did not differ between switching to Decision and switching to Prediction games.

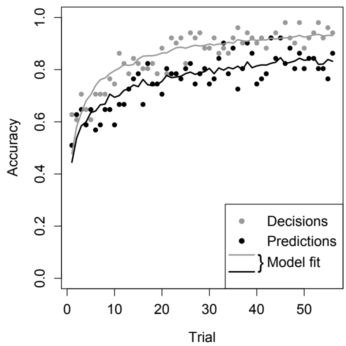

Figure B.1. Mean accuracy of decisions (orange) and predictions (blue), averaged across participants. The solid lines represent the fits of the linear mixed-effects models.

Figure B.1 shows that the accuracy of predictions is lower than the accuracy of decisions. This difference is significant, $\chi^2(1) = 44.77$, $p < .001$. At first, the probability of making correct decisions and predictions does not significantly differ, $\beta_{\text{prediction-decision}} = -.09$ (SE = .25), z = -.37, ns. However, the difference becomes larger, as can be seen in Figure B.1. At the end of the experiment, the probability of making a correct prediction is significantly lower than the probability of making a correct decision, $\beta_{\text{prediction-decision}} = -1.06$ (SE = .15), z = -7.10, $p < 0.001$.

Figure B.1 also shows that the accuracy of decisions and predictions increase with each game (i.e., trial) played, which is a significant main effect, $\chi^2(1) = 289.65$, $p < 0.001$. This trend shows that the participants became better as they played more games. However, the trend was smaller in prediction games than in decision games, $\beta_{log\text{-}trial} = 0.58$ (SE = 0.05), z = 10.78, and $p < 0.001$, and $\beta_{log\text{-}trial} = 0.82$ (SE = 0.06), z = 13.75, and $p < 0.001$, respectively. This interaction is significant, $\chi^2(1) = 9.11$, $p < 0.005$. Thus, performance was susceptible to improvement, but more so in Decision games than in Prediction games.

Response times

Participants' response (decision / prediction) times were also analyzed by means of linear mixed-effects models. The reaction times were first log-transformed to reduce skew in the distribution of response times. Performing the procedure of model comparison described above, we found a best fitting model that contained main effects, only, of ToM *order*, *switching*, and *trial*.

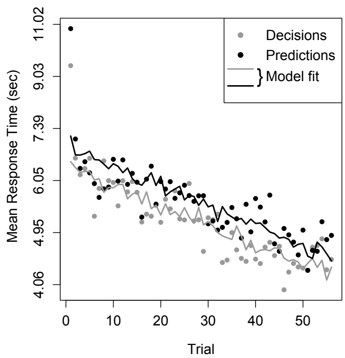

Figure B.2: The response times were averaged across participants; plotted separately for Decision games (orange) and Prediction games (blue) on a logarithmic scale.

The participants did not only make more mistakes in Predictions games than in Decision games (Figure B.1), they also required more time to predict the other player's decision than to make the same decision themselves (Figure B.2). This main effect is significant, $\chi^2(1) = 4.97$, $p < .05$.

Figure B.2 shows that the log-transformed response times (log-RTs) decreased linearly during the experiment, in both the Decision and Prediction games. This effect of *trial* on the log-RTs is significant, $\chi^2(1) = 354.01$, $p < .001$. The lack of an interaction between type of game (Decision / Prediction) and trial implies that the ratio of the actual decision and prediction times did not change over trial.

Interestingly, whereas switching between Decision and Prediction games did not have an effect on the accuracy of the participants' responses, switching did cause a significant time cost, $\chi^2(1) = 49.13$, $p < .001$. This finding implies that switching back and forth between one's own and the other player's perspective did have an associated time cost, in addition to the time cost associated with playing Prediction games instead of Decision games. Importantly, there were no interaction effects that included this switching factor, which means that the time cost of switching between Decision and Prediction games, and vice versa, was symmetrical. Therefore, switching between Decision and Prediction games could not have been a confounding factor in explaining, for example, shorter RTs in Decision games in comparison to Prediction games. That difference is solely to be attributed to the required order of theory of mind.