NeuroImage 137 (2016) 86–96

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Distinct neural patterns of social cognition for cooperation versus competition

Lily Tsoi [a,*], James Dungan [a], Adam Waytz [b], Liane Young [a]

[a] Department of Psychology, Boston College, Chestnut Hill, MA 02467, United States
[b] Department of Management and Organizations, Kellogg School of Management at Northwestern University, Evanston, IL 60208, United States

ARTICLE INFO

Article history:
Received 31 January 2016
Revised 12 April 2016
Accepted 29 April 2016
Available online 7 May 2016

Keywords:
fMRI
Social neuroscience
Theory of mind
Cooperation
Competition

ABSTRACT

How do people consider other minds during cooperation versus competition? Some accounts predict that theory of mind (ToM) is recruited more for cooperation versus competition or competition versus cooperation, whereas other accounts predict similar recruitment across these two contexts. The present fMRI study examined activity in brain regions for ToM (bilateral temporoparietal junction, precuneus, dorsomedial prefrontal cortex) across cooperative and competitive interactions with the same individual within the same paradigm. Although univariate analyses revealed that ToM regions overall were recruited similarly across interaction contexts, multivariate pattern analyses revealed that these regions nevertheless encoded information separating cooperation from competition. Specifically, ToM regions encoded differences between cooperation and competition when people believed the outcome was determined by their and their partner's choices but not when the computer determined the outcome. We propose that, when people are motivated to consider others' mental states, ToM regions encode different aspects of mental states during cooperation versus competition. Given the role of these regions for ToM, these findings reveal distinct patterns of social cognition for distinct motivational contexts.

Introduction

The capacity to attribute and reason about mental states, known as theory of mind (ToM), is no doubt critical for interpersonal interaction. An extensive body of prior work indicates a key role for theory of mind in moral evaluation (Ames and Fiske, 2013; Cushman, 2008, 2015; Decety and Cacioppo, 2012; Koster-Hale et al., 2013; Moran et al., 2011; Young et al., 2007, 2010; Young and Saxe, 2009), a social cognitive process essential for determining potential allies and enemies (Young and Waytz, 2013). An outstanding question, however, is whether people engage in ToM differently across different social motivational contexts. We suggest that successfully interacting with friends and foes depends on effective reasoning about their mental states (e.g., intentions, motivations, beliefs). For example, in chess, quickly taking the queen without considering why the other player sacrificed it could be perilous. Likewise, in the game *Taboo*, referencing obscure songs from the 1970s to clue in one's 10-year-old partner could bring down the team.

The current study uses functional magnetic resonance imaging (fMRI) to examine how people engage in ToM across two fundamental social contexts: cooperation versus competition. In the present research, we operationalized cooperation and competition according to goals and rewards. In cooperative contexts, interaction partners have the same goal, and if they achieve their shared goal, they both earn a reward. In competitive contexts, interaction partners have opposing goals, and only one individual can win in a zero-sum fashion. The present research targets cooperative and competitive interactions with the same individual, thereby limiting the influence of external social factors (e.g., group membership, familiarity, liking) and capturing processes common in everyday life (e.g., two friends who try to outpace each other in their tasks at work might also get together to make dinner afterwards).

Separate lines of research on the evolutionary origins of ToM, intergroup cognition, and motivational relevance provide different perspectives on the broad question of whether people engage in ToM differently for cooperation versus competition. In one line of research on the evolutionary origins of ToM, rudimentary ToM capacities are observed among non-human primates but only in the evolutionarily and ecologically salient domain of competition (e.g., fighting over scarce resources such as food) (Hare, 2001; Hare and Tomasello, 2004; Lyons and Santos, 2006). Even though some work has revealed successful acts of cooperation among some non-human primate species (Tomasello and Vaish, 2013), ToM capacities do not appear to extend to cooperative or collaborative contexts to the same extent as for competitive contexts (Lyons

Abbreviations: dmPFC, dorsomedial prefrontal cortex; lTPJ, left temporoparietal junction; MVPA, multivariate pattern analyses; rTPJ, right temporoparietal junction; ROI, region of interest; ToM, theory of mind.

* Corresponding author at: 140 Commonwealth Avenue, McGuinn 300, Chestnut Hill, MA 02467, United States.
E-mail address: lily.tsoi@bc.edu (L. Tsoi).

http://dx.doi.org/10.1016/j.neuroimage.2016.04.069
1053-8119/

and Santos, 2006; but see Schmelz and Call, 2016). The primarily competitive nature of social interactions among non-human primates and selective pressures such as limited availability of resources (e.g., necessities such as food) may have favored individuals with the ability to represent the intentions, perceptions, and simple beliefs of other creatures. This selective advantage may have persisted in humans. Indeed, research in humans provides some evidence for greater deployment of ToM for competition than cooperation. Neuroimaging research (discussed later in connection to the present findings) provides some initial evidence of differential ToM processing for competitive versus cooperative contexts (Decety et al., 2004; Lissek et al., 2008), and behavioral studies in human adults and children show that an agent's negative behavior, as compared to neutral or positive behavior, is a particularly strong trigger for ToM in the service of understanding the agent's present and future behaviors (Morewedge, 2009; Vaish et al., 2008; Waytz et al., 2010a). Negative stimuli, as compared to positive stimuli, are more attentionally salient (e.g., elicit longer looking times), perceived as more complex, and thought to carry more weight or greater informational value (Fiske, 1980; Peeters and Czapinski, 1990; Vaish et al., 2008). People's attempt to gain control over a negative agent may lead to this asymmetry, resulting in more complex cognitive representations of negative stimuli (Peeters and Czapinski, 1990).

A different line of research on intergroup processes predicts greater ToM for cooperation versus competition. This work reveals that people consider the minds of others differently depending on group membership. That is, people tend to attribute mental states more comprehensively to ingroup members than to outgroup members (Kelman, 1973; Leyens et al., 2000; Opotow, 1990; Struch and Schwartz, 1989). Some accounts suggest that actual or perceived competition over limited resources (e.g., food, money, jobs) with outgroup members is what drives intergroup conflict and hostility (realistic group conflict theory; Jackson, 1993), leading people to disregard the emotional experience of outgroup members (Cikara et al., 2011). This tendency, known as dehumanization, has been observed explicitly and implicitly (Haslam and Loughnan, 2014) as well as across different target outgroups pertaining to ethnicity, race, gender, and disability (Haslam, 2006). Consistent with these behavioral findings is neural evidence showing that evaluating others extremely dissimilar to the self compared to similar others fails to elicit activity in the medial prefrontal cortex, a brain region implicated in social cognition and mentalizing (Harris and Fiske, 2006). Given separate work showing that people tend to cooperate more with ingroup members (McAuliffe and Dunham, 2016), consider ingroup minds more than outgroup minds (based on the dehumanization literature), and attend more to the minds of those with whom they wish to affiliate and cooperate (Kozak et al., 2006), we might predict greater ToM for interaction partners during affiliative and cooperative interactions. This pattern might persist even when—as in the present study—the same person is sometimes a competitor and other times a cooperator.

Another line of work offers a different interpretation: ingroup members may be generally more motivationally relevant than outgroup members, leading people to attend more to ingroup minds than to outgroup minds. One prediction of this motivational account is that people engage in ToM when they are motivated by cooperative and competitive goals regardless of group membership. Some social psychological work supports this idea: people have more lenient thresholds for perceiving minds behind ingroup versus outgroup faces, but this pattern changes for outgroup faces that are perceived as threatening (Hackel et al., 2014). More generally, how people perceive and interact with others (ingroups versus outgroups, cooperators versus competitors) may depend on motivational factors such as effectance motivation and affiliation motivation (Epley et al., 2007; Waytz et al., 2010b; White, 1959). When interacting with others, people may be guided by the motivation to predict others' actions and gain mastery over their environment (effectance motivation), or, by contrast, people may be guided by the desire for social contact and affiliation (affiliation motivation). Research shows that these two different types of motivation lead to preferential focus on two different dimensions of mind perception (Waytz and Young, 2014): agency (i.e., capacity for planning, thinking, intending) and experience (i.e., capacity for emotion, feeling) (Gray et al., 2007; Gray and Wegner, 2009). Specifically, effectance motivation and affiliation motivation lead people to focus more on agentive and experiential mental states, respectively. While effectance motivation might help people plan attacks and outsmart enemies in competitive interactions, affiliation motivation might help people create and maintain alliances during cooperative interactions. This prior work therefore predicts that people may engage in ToM for cooperative *and* competitive interactions; understanding others' minds may help people respond appropriately in both types of social contexts though the dimensions of mind perception and underlying motivations may differ. Therefore, along with investigating whether brain regions implicated in ToM are recruited similarly robustly for both cooperative and competitive interactions, the present work also examines whether ToM regions encode, in their spatial pattern of activity, any difference in how participants process the mental states of the same individual depending on whether that individual is a cooperator or competitor.

Neuroimaging work has revealed a network of brain regions that is reliably and robustly recruited when people engage in ToM (Fletcher et al., 1995; Gallagher et al., 2000; Gobbini et al., 2007; Saxe and Kanwisher, 2003), including for moral judgment (Young et al., 2007; Young and Saxe, 2008, 2009). These regions include right and left temporoparietal junction (rTPJ and lTPJ), precuneus, and dorsomedial prefrontal cortex (dmPFC). We extend the current research on ToM by using fMRI to examine whether and how these ToM regions are recruited for social interactions that differ only in terms of whether they are cooperative or competitive. Not many studies have directly compared ToM for cooperation with ToM for competition, though prior neuroimaging research has separately revealed recruitment of ToM regions during cooperative situations or social games assessing cooperative intent (Elliott et al., 2006; Krueger et al., 2007; McCabe et al., 2001; Rilling et al., 2004) as well as recruitment of ToM regions during competitive situations (Gallagher et al., 2002; Hampton et al., 2008). The present study directly compares cooperative with competitive interactions involving the same person within the same paradigm, using activity in the ToM network as a proxy for the cognitive process of ToM.

The current study has two main goals: (1) to investigate the overall recruitment of ToM regions for cooperative and competitive interactions and (2) to examine whether ToM regions encode, in their spatial pattern of activity, information separating cooperative from competitive interactions. Because prior work has uncovered the capacity of simple economic games to recruit brain regions for ToM (Krueger et al., 2007; McCabe et al., 2001; Rilling et al., 2004), we decided to use an a similar methodological approach in the current study. We designed a novel dyadic game "Shapes" modeled after "Rock, Paper, Scissors". Cooperation and competition are operationalized in terms of goals (shared vs. opposing goals, respectively) and payoffs (wins and losses are yoked vs. one person wins and one person loses, respectively). In *active* trials, trial outcomes depend on both players' responses. In *passive* trials, trial outcomes are determined by the computer. By including *passive* conditions, we can better identify constraints on the information encoded by ToM regions. If ToM regions are sensitive simply to goal-oriented differences or to payoff-oriented differences, the spatial patterns of neural activity in ToM regions for cooperative and competitive interactions may be distinct from each other in *both* active and passive conditions. If ToM regions are sensitive to differences between cooperative and competitive contexts for active but not passive conditions, ToM regions may encode information about the opponent's (competitor's) or ally's (cooperator's) mental states but only when those mental states guide behavior relevant to the interaction (e.g., when players' behaviors are thought to determine outcomes). To test whether ToM is preferentially engaged for cooperation versus competition, we use univariate analyses to examine whether the magnitude of activity in ToM regions

differs for cooperative and competitive interactions. To test whether the representational content of ToM regions differs for cooperation versus competition, we use multivariate pattern analyses to examine whether the spatial patterns of activity within ToM regions are different for cooperative and competitive interactions.

Materials and methods

Participants

Nineteen right-handed participants between the ages of 21 and 38 (mean: 27.16 ± 5.11; 8 females) were recruited from the Boston area. All participants were native English speakers, had normal or corrected-to-normal vision, and reported no history of psychiatric or neurological disorders. Participants gave written informed consent and were paid $25/h for their participation plus a $36 bonus (see below for details) for their participation in the game. The study was approved by the Boston College Institutional Review Board.

One participant was excluded from all subsequent analyses due to excessive motion (e.g., within-run motion > 6 mm). The final set of participants in the analyses of the neural and in-scanner behavioral data consisted of 18 adults (8 females, ages 21–38, $M = 27.4 \pm 5.1$).

Experimental task

During the consenting process, the participant was introduced to a gender-matched confederate, ostensibly another participant. In reality, the participant interacted with a computer program for the entire duration of the experimental task (see Inline Supplementary Text and Inline Supplementary Fig. S1 for comparisons between this task and an actual two-player version of the task conducted in the lab; Inline Supplementary Fig. S1 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069.). Both the participant and the confederate were told that the experimental task involved two players interacting with each other via a computer interface, during which one person would be scanned first, and the other person would be scanned second. The participant and the confederate drew straws to determine the scanning order. The straws were fixed such that the participant (and not the confederate) would be scanned first (the confederate was not scanned; participants were fully debriefed shortly after the scan session).

The experimental task was presented to participants as a game variant of "Rock, Paper, Scissors" with two shapes (i.e., circle or square). The task had a 2 (context: cooperative, competitive) × 2 (goal: active, passive) within-subjects design (Fig. 1). In *active* conditions, both players were given a goal: to either match (i.e., choose the same shape as) or mismatch (i.e., choose the shape opposite of) the other player's shape, and the outcome of the trial depended on their responses. In *active cooperative* trials, both players were given the same goal, and both players won $1 if they fulfilled their shared goal (e.g., if both players' goals were to match each other's shape and they both chose "circle", then both would win $1; if both players' goals were to mismatch the other's shape and one person chose "circle" while the other chose "square", they both would win $1). In *active competitive* trials, players had opposing goals, and only one person would be able to fulfill their goal and win $1 (e.g., if the participant's goal was to match while the other player's goal was to mismatch and both players choose "square", only the participant would win $1; if the participant's goal was to mismatch while the other player's goal was to match and both players chose "square", this time the other player would win $1 and the participant would win nothing).

In *passive* conditions, both players were assigned a shape, and the outcome of each trial depended on the computer's "randomly" generated shape (this was in fact fixed). In *passive cooperative* trials, both players were assigned the same shape, and both could win $1 depending on the shape the computer "randomly" generated for that trial; in *passive competitive* trials, players had opposing shapes, and only one person could win $1 based on the computer's "randomly" generated shape. Thus, the payoff structure from cooperative and competitive *passive* conditions mimicked that of cooperative and competitive trials from *active* conditions, respectively.

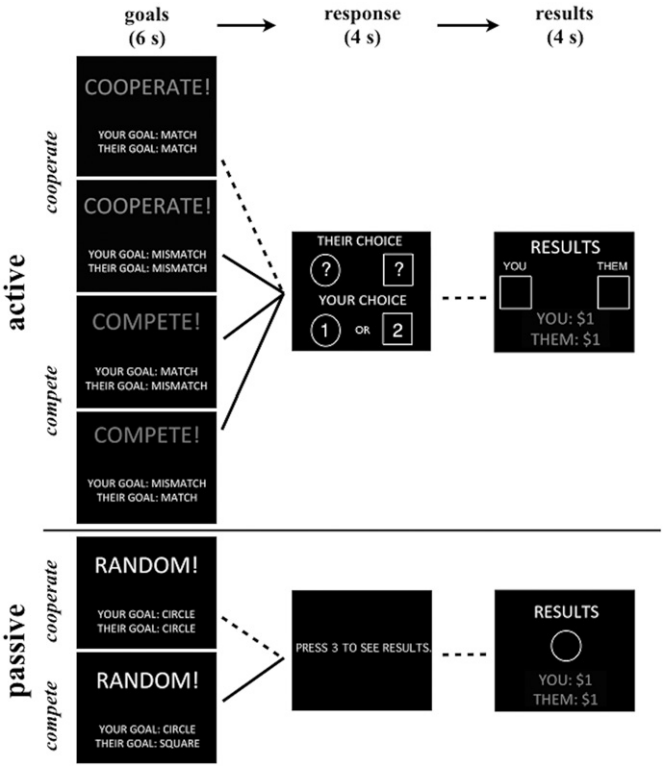

Fig. 1. Experimental task. In each trial, participants are shown both players' goals (goals phase), are told to make their responses within a specified time window (response phase), and are given feedback about the trial (results phase). Possible trials from the active condition and passive condition (in dashed lines) are shown; there are six conditions in total.

The results of the experimental task were fixed: every participant won half of the trials and lost half of the trials (again, see Inline Supplementary Text and Inline Supplementary Fig. S1 for comparisons between this task and an actual two-player version of the task). At the end of the experiment, participants received a bonus of $36 ($1 for each trial, for 36 out of 72 trials).

Each trial consisted of three phases (Fig. 1): the *goals* phase (6 s), which provided all the necessary information (i.e., both players' goals) for the participant to generate a response; the *response* phase (4 s), during which participants made their responses (i.e., either by pressing one button for "circle" or a second button for "square" in active conditions, and by pressing a third button in passive conditions); and the *results* phase (4 s), which provided feedback for each trial (i.e., the shapes that both players selected or the shape that the computer "randomly" generated as well as the amount each player earned for that trial). A blank screen (resting period) was presented between trials (12 s). There were six 5-minute runs in total; two trials in each of the six trial types (active-match cooperative; active-mismatch cooperative; active-match competitive; active-mismatch competitive; passive-cooperative; passive-competitive) were presented in each run (one participant completed only four runs). The order in which trials were presented was pseudorandom.

After the scan session, participants were asked to fill out a demographics questionnaire, a post-game survey probing their reactions during the game (e.g., how much participants were likely to befriend the other player; the extent to which participants felt like they were cooperating and competing during cooperative and competitive trials, respectively), the Autism Spectrum Quotient, and the Interpersonal

Reactivity Index (see Inline Supplementary Table S1, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069). Because no correlations were found between any of these measures and neural data (mean response magnitude or classification accuracy in ToM regions), these measures are not discussed in the paper.

After participants were debriefed, 5 out of the 18 participants expressed occasional doubt that they were playing with an actual person during parts of the game; when probed further, these participants reported that their occasional doubt did not alter their behavior in the game. We compared the behavioral responses, reaction times, and neural data for participants who expressed doubt with those who did not express any doubt and found no group differences. Because we did not find any group differences, we included all 18 participants in our analyses.

Imaging procedure

fMRI data acquisition

Participants were scanned at the Harvard Center for Brain Science. Anatomical and functional data were acquired using a Siemens 3.0 T Tim Trio MRI scanner and a 12-channel head coil. Thirty-six axial slices (3-mm isotropic voxels, 0.54-mm gap) were acquired using the following gradient-echo planar imaging (EPI) sequence parameters: repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle (FA) = 90°; field of view (FOV): 216 × 216; interleaved acquisition. Stimuli were generated on an Apple MacBook Pro running MATLAB 2008b with Psychophysics Toolbox. Stimuli were projected onto a screen (1024 × 768 pixel resolution) at the end of the magnet bore, which participants viewed via a mirror mounted on the head coil.

fMRI data preprocessing

Data preprocessing and analyses were performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm) and custom software. Functional data were corrected for slice timing, realigned to the first EPI, spatially normalized onto a common brain space (Montreal Neurological Institute, MNI), spatially smoothed using a Gaussian filter (full-width half-maximum = 8 mm kernel), and high-pass filtered (128 Hz).

fMRI analysis

For the experimental task, a slow event-related design was used and modeled using a boxcar regressor convolved with a canonical hemodynamic response function (HRF). The boxcar function extended for the entire 14-s period (7 TRs) during which the goals phase, response phase, and results phase were presented. The general linear model (GLM) included movement parameters as nuisance regressors.

Whole-brain univariate analysis. Beta values were estimated in each voxel for all six conditions (cooperate-match, cooperate-mismatch, compete-match, compete-mismatch, cooperate-passive, compete-passive). Four contrast maps were produced in each subject identifying voxels responding more to: 1) $coop_{active} > coop_{passive}$, 2) $comp_{active} > comp_{passive}$, 3) $coop_{active} > comp_{active}$, and 4) $comp_{active} > coop_{active}$. Contrast images for each comparison were submitted to a random effects analysis. To correct for multiple comparisons, contrast images from random effects analyses were subjected to a voxel-wise threshold of $p < 0.001$ (uncorrected) and a cluster extent threshold ensuring a FWE rate of $p < 0.05$ ($k = 14$ voxels), based on cluster extent and Gaussian random field theory as implemented in SPM8 (Friston et al., 1994; Worsley et al., 1992). Anatomical labels for peak coordinates were retrieved using SPM Anatomy Toolbox v1.8 (Eickhoff et al., 2005).

Defining ROIs. A ToM localizer task (Dodell-Feder et al., 2011) was used to functionally define the following regions for ToM: the rTPJ, lTPJ, precuneus, and dmPFC. The task consisted of 10 stories in each of two conditions: (1) stories requiring the inference of another person's mental states (e.g., false beliefs) and (2) stories requiring the inference of outdated (i.e., false) physical representations (e.g., outdated photographs). The entire set of stimuli can be found at: http://saxelab.mit.edu/superloc.php. Each story was presented on the screen for 10 s, followed by a true/false question about the story (4 s). An event was defined as the period between the start of the story presentation and the end of the question presentation (14 s). Beta values were estimated in each voxel for stories describing mental states (e.g., belief) or physical representations (e.g., photo). A contrast map was produced in each subject identifying voxels responding more to stories about beliefs than stories about photos. ROIs were defined as all voxels in a 9-mm radius of the peak voxel that passed threshold in the contrast image belief > photo ($p < 0.001$, uncorrected; $k > 16$, a value computed via 1000 iterations of a Monte Carlo simulation) (Slotnick et al., 2003). For mean peak coordinates for each ROI, see Inline Supplementary Table S2, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069.

ROI-based univariate analysis. The neural response over baseline to each condition was calculated for each ROI. Baseline response in each ROI was calculated as the average response in that ROI at all time points during the resting period, excluding the first 6 s after the offset of each stimulus (to allow the hemodynamic response to decay). The percent signal change (PSC) relative to baseline was calculated for each time point in each condition, averaging across all voxels in the ROI, where PSC (at time t) = 100 × [(average magnitude response for condition at time t − average magnitude response for fixation) / average magnitude response for fixation]. PSC was averaged across the entire trial (7 TRs or 14 s; offset 6 s from presentation time to adjust for hemodynamic lag) to estimate a single PSC for each condition in each ROI in each participant.

ROI-based multivariate analysis. fMRI time courses from voxels within ROIs were extracted from unsmoothed data and high-pass filtered (128 Hz) in order to remove slow drifts. Each voxel's time course was zero-meaned. There were four stimulus types of interest ($coop_{active}$, $comp_{active}$, $coop_{passive}$, and $comp_{passive}$) and two comparisons of interest ($coop_{active}$ versus $comp_{active}$ and $coop_{passive}$ versus $comp_{passive}$). The TRs corresponding to each of the six conditions were labeled as one of the four stimulus types. For each stimulus type, a regressor was constructed by convolving the onset of each trial of that stimulus type with the canonical HRF (hemodynamic response function). The mean of the height of the regressor was calculated, and each time point was assigned to that stimulus type if the height of the regressor at that time point was greater than the mean height. For each comparison, a binary classification was performed in each ROI using a Gaussian Naive Bayes (GNB) classifier (Pereira et al., 2009; Raizada and Lee, 2013). To train the classifier, a subset of the data (5 out of 6 runs) was used to build a model that appropriately set the boundary between neural activity associated with one stimulus type and neural activity associated with the other stimulus type in the comparison of interest. This model was applied to the remaining data (1 out of 6 runs) for validation. Accuracy of the classification test was calculated by calculating the number of times the classifier correctly predicted the time points corresponding to the conditions being compared. Leave-one-run-out cross-validation was used for all analyses. An accuracy score averaged across training/testing set combinations was computed for each individual and each ROI.

Results

In-scanner behavioral results

Although our main analyses targeted neural data, we conducted an exploratory analysis of behavioral data collected in the scanner. In particular, we tested whether participants relied on different strategies for each of four possible types of active trials (cooperate-match, cooperate-mismatch, compete-match, compete-mismatch). Within each trial type,

we examined participants' patterns of behavioral responses for each trial in relation to the previous trial. We specifically examined how participants' responses matched the following strategies: (1) choosing the same shape they had chosen in the previous trial (shape-self), (2) choosing the shape the other player had chosen on the previous trial (shape-other), and (3) following a win-stay lose-shift (WSLS) strategy (if they won on the previous trial, then stay with the same shape, otherwise switch shape). We calculated the proportion of cooperative or competitive trials that matched the WSLS template, the proportion of trials that matched the Shape-self template, and the proportion of trials that matched the Shape-other template. Note that a participant's response on a particular trial could potentially match one, two, or all three of the specified strategies (see Inline Supplementary Table S4 and Inline Supplementary Fig. S2, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069). We chose these strategies because of their close correspondence to strategies often described in evolutionary game theory (Axelrod and Hamilton, 1981; Imhof et al., 2007; Nowak and Sigmund, 1993).

We performed a 2 (context: cooperate, compete) × 2 (goal type: match, mismatch) × 3 (strategy: shape-self, shape-other, WSLS) repeated-measures ANOVA. Although there was no three-way interaction, there was a significant context by strategy interaction, $F(2, 34) = 7.194$, $p = 0.002$, indicating that participants relied on different strategies depending on the context (Fig. 2; see Inline Supplementary Table S3 and Inline Supplementary Table S4, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069). Pairwise comparisons showed that the percentage of cooperative trials that matched the shape-self strategy was significantly higher than that for shape-other ($t(17) = 4.387$, $p < 0.001$) and WSLS ($t(17) = 4.622$, $p < 0.001$), with no significant difference between shape-other and WSLS ($t(17) = -.513$, $p = 0.615$). By contrast, there was no single strategy for competitive trials; the percentage of competitive trials was similar for all three strategies ($ps > 0.05$). Overall, these results suggest that participants deployed different strategies for cooperative and competitive trials. Participants were more likely to stick with the same shape for cooperative trials, perhaps as a way to signal their choice to the other player. By contrast, participants' responses in competitive trials did not appear to follow a straightforward strategy. On competitive trials, participants may have been responding in a more unpredictable or random manner with the aim of confusing the other player. This pattern is consistent with prior work showing that people are less willing to be predicted in competitive situations compared to cooperative situations (Ybarra et al., 2010).

Additionally, we tested for changes of strategies and/or behaviors over time. We examined whether the extent to which participants' behaviors matched each strategy changed over time. The proportion of trials that matched each of the three strategies (i.e., Shape-self, Shape-other, WSLS) did not differ for the first and second halves of the task ($ps > 0.05$). We also examined participants' patterns of responses based on their goals and responses on the immediately preceding trial. We counted (1) the number of times a participant's response (i.e., circle or square) stayed the same when the participant's goal was the same (e.g., trial 1: match; trial 2: match) and (2) the number of times a participant's response changed when the goal changed (e.g., trial 1: match; trial 2: mismatch). We compared the first half of the task to the second half of the task for these two sets of values. We found that the number of times a participant chose the same shape as the preceding trial when his/her goal was the same as that of the preceding trial did not differ for the first and second halves of the task (first half: $M = 2.5$, $SD = 1.62$; second half: $M = 2.78$, $SD = 1.48$; $p = 0.60$). The number of times a participant changed his/her shape when his/her goal was different from that of the preceding trial also did not differ for the first and second half of the task (first half: $M = 4.11$, $SD = 1.75$; second half: $M = 3.5$, $SD = 1.62$; $p = 0.21$). Note that these values were low overall, making it difficult to divide the data further by context. Overall, these results suggest that participants' strategies and behavioral patterns did not change systematically over the duration of the session.

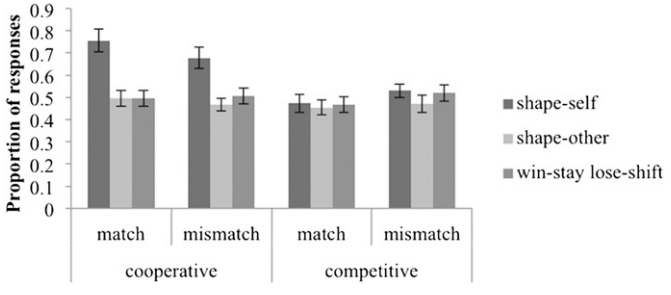

Fig. 2. Proportion of responses for each active condition that matched each of three different strategies. Strategies included shape-self (choosing the same shape chosen on the previous trial of the same condition), shape-other (choosing the same shape as the other player on the previous trial of the same condition), and win-stay lose-shift (choosing the same shape chosen on the previous trial of the same condition if participant won on that trial; otherwise, choose the other shape). Each response may match multiple strategies (see Inline Supplementary Fig. S2; see Inline Supplementary Table S4). Error bars denote SEM.

Analyses involving reaction time data can be found in Supplementary materials, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069.

fMRI results

ToM regions were recruited for active versus passive trials

Whole-brain univariate analyses revealed the following regions for *active versus passive cooperative* trials ($coop_{active} > coop_{passive}$): calcarine gyrus, right cuneus, left superior medial gyrus, right angular gyrus, left angular gyrus, left superior occipital gyrus, middle frontal gyrus, and right cerebellum (Table 1). Regions recruited more for *active versus passive competitive* trials ($comp_{active} > comp_{passive}$) included the left insula lobule, left middle orbital gyrus, right cuneus, right supramarginal gyrus, right inferior frontal gyrus, right middle orbital gyrus, left cuneus, left angular gyrus, left superior medial gyrus, right fusiform gyrus, and right precuneus (Table 1). For whole-brain univariate analyses for *active versus passive* trials (*active* > *passive*), see Inline Supplementary Table S5. For ROI univariate analyses for *active versus passive* trials, see Inline Supplementary Table S6.

Inline Supplementary Tables S5 and S6 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069.

We performed separate conjunction analyses examining overlap in regions for the *belief* > *photo* contrast from the ToM localizer task and (1) regions for the $coop_{active} > coop_{passive}$ contrast and (2) regions for the $comp_{active} > comp_{passive}$ contrast. The first conjunction analysis revealed the rTPJ, lTPJ, dmPFC (Fig. 3A); the second conjunction analysis revealed the rTPJ, lTPJ, and precuneus (Fig. 3B). These results show that regions recruited for active versus passive interactions during both cooperative and competitive trials overlapped with ToM regions as elicited by our independent localizer task. A conjunction analysis of all three contrasts also revealed the ToM network (see Inline Supplementary Fig. S3, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069). To correct for multiple comparisons, each contrast image was subjected to a cluster extent threshold ensuring a FWE rate of $p < 0.05$.

ToM regions overall responded similarly to cooperative and competitive interactions

We restricted our analyses to our conditions of interest: active cooperative and active competitive trials collapsing across match and mismatch trials. Whole-brain univariate analyses revealed no widespread recruitment of regions implicated in ToM for active cooperative versus active competitive trials or vice versa (cluster-level corrected at $p < 0.05$). Instead, the $coop_{active} > comp_{active}$ contrast revealed the right

Table 1
Results from whole-brain univariate analyses.

Region name	x	y	z	*t* value	# voxels
Cooperative: active > passive					
Calcarine gyrus	0	−88	−8	8.31	44
R cuneus	12	−97	16	7.49	151
L superior medial gyrus	3	41	37	7.38	409
R angular gyrus	48	−52	37	6.68	248
L angular gyrus	−45	−58	28	6.02	109
L superior occipital gyrus	−12	−103	13	5.81	164
L middle frontal gyrus	−39	23	40	5.81	87
R cerebellum	36	−70	−29	5.13	122
Competitive: active > passive					
L insula lobule	−27	20	−14	8.07	80
L middle orbital gyrus	−39	47	−5	7.69	249
R cuneus	9	−100	16	7.54	101
R supramarginal gyrus	51	−46	34	7.31	280
R inferior frontal gyrus	36	26	−14	7.21	89
R middle orbital gyrus	36	59	−2	7.20	182
L cuneus	−9	−103	16	6.83	55
L angular gyrus	−51	−55	31	6.57	185
L superior medial gyrus	3	32	37	6.48	129
R fusiform gyrus	30	−73	−17	6.41	490
R precuneus	3	−67	43	4.81	43
Cooperative > competitive (active)					
R supplementary motor area	3	−7	52	6.28	67
Competitive > cooperative (active)					
Supplementary motor area	0	20	46	7.10	86
L inferior frontal gyrus	−48	20	31	6.58	49
R precuneus	9	−67	52	5.81	113

Note: Different contrasts are bolded and italicized. All regions listed survived cluster-level correction ($p < 0.05$). All coordinates are in MNI.

supplementary motor area, while the $comp_{active} > coop_{active}$ contrast revealed a more anterior region of the supplementary motor area, the left inferior frontal gyrus, and the right precuneus (Table 1). Even when we applied no cluster-level correction and examined only the results using a voxel-wise threshold of $p < 0.001$ (uncorrected) with an extent threshold of 10 voxels, the contrasts did not reveal the ToM network.

Next, taking a conservative approach, we conducted a series of ROI-based univariate analyses of responses averaged over the entire time course (from the beginning of the goals phase to the end of the response phase; 7 TRs). Because we wished to focus on the effect of context (cooperative, competitive), we also collapsed across goal type (match, mismatch) in all subsequent analyses; the effect of context (cooperative, competitive) did not differ for match and mismatch trials in most ROIs ($ps > 0.05$, except for lTPJ; $F(1, 15) = 5.894$, $p = 0.028$).

To hone in on whether ToM regions were differentially recruited for active cooperative and active competitive trials, we compared neural activity for active cooperative conditions with neural activity for active competitive conditions using ROI-based univariate analyses (Fig. 4A). We performed a 2 (context: active cooperative, active competitive) × 4 (ROI: rTPJ, lTPJ, precuneus, dmPFC) repeated-measures ANOVA with a Greenhouse–Geisser correction, which revealed a marginal interaction ($F(3, 36) = 3.047$, $p = 0.062$) and, critically, no main effect of context ($F(1, 12) = 2.028$, $p = 0.180$). Pairwise comparisons (uncorrected for multiple comparisons) for each ROI revealed no difference in the response magnitude for active cooperative versus active competitive trials for the rTPJ ($t(16) = -1.519$, $p = 0.148$), precuneus ($t(14) = -.749$, $p = 0.466$), and dmPFC ($t(14) = 1.652$, $p = 0.121$), and a marginal difference for the lTPJ ($t(15) = -2.124$, $p = 0.051$) (note: because we were unable to define all ROIs for all participants using the localizer task, degrees of freedom differ across ROIs). A similar pattern (i.e., no difference for most ToM regions) held for wins and losses separately (see Inline Supplementary Table S7, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069). Together, these analyses suggest that most ToM regions respond similarly to active cooperative and active competitive trials. Similar patterns were found for passive trials (see Inline Supplementary Fig. S4, which can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069); that is, ROI-based univariate analyses revealed that neural responses for passive cooperative and competitive trials were not significantly different from each other in any ROI ($ps > 0.05$).

We note that similar patterns were found across all three phases of the trial (i.e., goals, response, and results). To address concerns that independence assumptions were violated with regard to phase, we used a mixed models approach instead of a repeated-measures ANOVA. We performed a linear mixed effects analysis: as the Y variable, we used percent signal change for each phase; as fixed effects, we entered context, phase, and the interaction term into the model; as random effects, we had intercepts for subjects and by-subject random slopes for the effect of phase. We compared models with and without an autoregressive covariance structure and selected the model without an autoregressive covariance structure based on the likelihood ratio test and Akaike's Information Criteria (AIC); thus, we report statistics for the simpler model. The interaction between context and phase was not significant in any ROI ($ps > 0.05$; lowest $p = 0.135$).

ToM regions encoded differences between cooperative and competitive trials in active but not passive conditions

Using multivariate pattern analyses (MVPA), we next examined whether the spatial patterns of neural activity for cooperative trials and competitive trials in ToM ROIs were distinct from each other. We tested whether the spatial patterns for cooperative trials and competitive trials within each ROI could be accurately classified above chance (50%) (Fig. 4B). Note that one-tailed one-sample t-tests were performed for each ROI; we were interested in examining whether classification

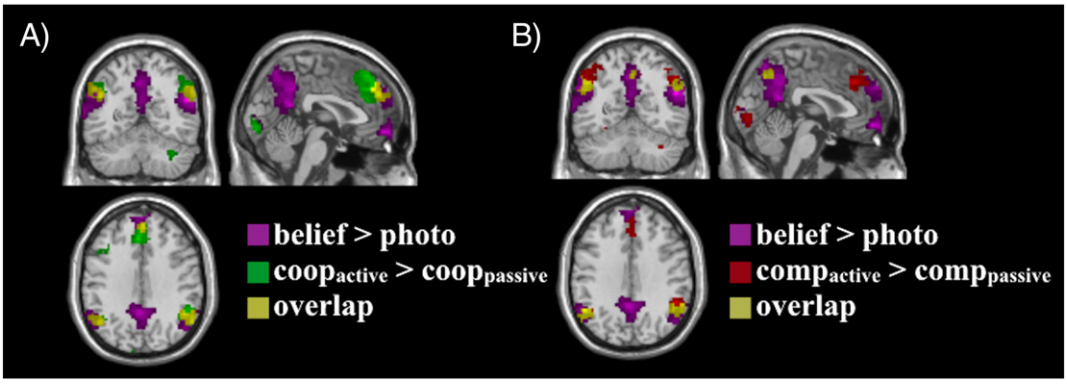

Fig. 3. Conjunction analyses. (A) Regions in the $coop_{active} > coop_{passive}$ contrast from the experimental task (green) overlapped with ToM regions in the *belief* > *photo* contrast from an independent localizer task (purple) [overlap in yellow]. (B) Regions in the $comp_{active} > comp_{passive}$ contrast from the experimental task (red) overlapped with ToM regions in the *belief* > *photo* contrast (purple) [overlap in yellow]. Each contrast image was subjected to a cluster extent threshold ensuring a FWE rate of $p < 0.05$.

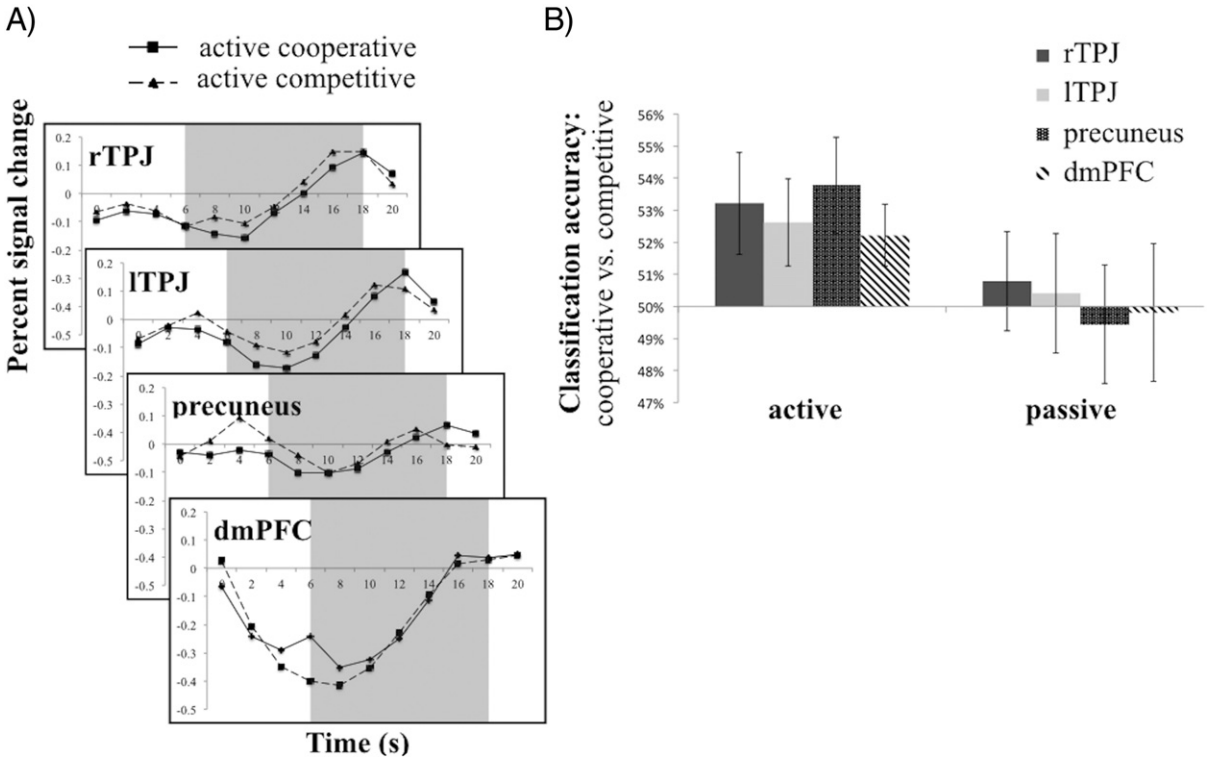

Fig. 4. Results from ROI-based univariate analyses (A) and multivariate pattern analyses (B). (A) Time courses for active cooperative and active competitive conditions for the following ROIs: rTPJ, lTPJ, precuneus, and dmPFC. Gray highlighted sections reflect the time window during which a trial was presented, adjusted for hemodynamic lag. (B) Mean classification accuracy in each ROI for cooperative and competitive trials for active conditions (significantly above chance in all ROIs; left) and passive conditions (at chance for all ROIs; right). Error bars denote SEM. See also Inline Supplementary Fig. S4; see Inline Supplementary Table S7.

accuracy is above chance (Mur et al., 2008) and not below chance (below-chance accuracies are not interpretable). Spatial patterns of activity for active cooperative trials and active competitive trials were classified above chance in all four ROIs (rTPJ: $M = 0.53 \pm 0.07$, $t(16) = 2.019$, $p = 0.030$; lTPJ: $M = 0.53 \pm 0.05$, $t(15) = 1.909$, $p = 0.038$; precuneus: $M = 0.54 \pm 0.06$, $t(14) = 2.511$, $p = 0.012$; dmPFC: $M = 0.52 \pm 0.04$, $t(14) = 2.313$, $p = 0.018$). By contrast, classification accuracy for passive cooperative trials and passive competitive trials was at chance (50%) for all four ROIs (all $ps > 0.05$). These results show above-chance discrimination between cooperative and competitive trials in active but not passive conditions across ToM ROIs. We note that paired-samples t-tests comparing classification accuracies for active and passive conditions revealed non-significant trends (ps were between 0.053 and 0.18 for all four ROIs). Combining information across ROIs (i.e., including voxels from all defined ROIs as features in the MVPA) led to similar results: at-chance classification for passive cooperative and competitive trials ($M = 0.506$, $SD = 0.084$) ($t(17) = 0.325$, $p = 0.37$) versus above-chance classification accuracy for active cooperative and competitive trials ($M = 0.540$, $SD = 0.058$) ($t(17) = 2.908$, $p = 0.0049$).

We also examined whether our above-chance classifications were driven by differences in behavioral strategies or reaction times for the two contexts (cooperative vs. competitive). To examine whether above-chance discrimination between active cooperative and competitive trials is driven by differences in behavioral strategies for the two contexts, we calculated 3 difference scores for each participant: % of trials matching a WSLS strategy for cooperation - % of trials matching a WSLS strategy for competition; % of trials matching a Shape-self strategy for cooperation - % of trials matching a Shape-self strategy for competition; and % of trials matching a Shape-other strategy for cooperation - % of trials matching a Shape-other strategy for competition. These difference scores measured the extent to which participants engaged in a strategy more so during one context than the other; higher absolute values indicated greater reliance on a specific strategy for one context versus the other. Classification accuracy did not correlate with any of the difference scores in any ToM ROI ($ps > 0.05$). To examine whether above-chance discrimination between cooperative and competitive trials is related to reaction time, we performed correlations between differences in reaction time for active cooperative and competitive trials and classification accuracy for active cooperative and competitive trials: the correlations were not significant in any ToM ROI ($ps > 0.05$). These findings suggest that discriminability between cooperation and competition within ToM regions is not related to differences in behavioral strategies or reaction time for the two contexts.

We then examined whether the information encoded by ToM regions is represented at a fine-grained scale or at a relatively coarser-grained scale. Prior work has investigated this question by testing how smoothing fMRI data affects decoding accuracies (Ethofer et al., 2009). The logic is as follows: decoding from large-scale representations would improve when fMRI data is smoothed because of increased signal-to-noise ratio within subregions, whereas decoding from fine-scale representations would get worse when fMRI data is smoothed because of decreased contrast between voxels (but see Kamitani and Sawahata, 2010). The ROI-based MVPA we performed were on unsmoothed data. When we performed ROI-based MVPA comparing active cooperative and active competitive trials using smoothed data with 8 mm FWHM, we found that mean classification accuracy was numerically above chance for all ROIs, though the effect was significant only for the precuneus, $t(14) = 2.08$, $p = 0.028$. The effect was marginal for rTPJ ($p = 0.090$), lTPJ ($p = 0.090$), and dmPFC ($p = 0.054$). These results suggest that, with smoothed data, discriminability between cooperation and competition dropped to levels not significantly different from chance in most ToM regions—tentatively indicating relatively fine-grained representation of information about context.

Discussion

One major goal of social neuroscience is to gain a greater understanding of collective human behavior, both positive (e.g., cooperation)

and negative (e.g., conflict and war) (Adolphs, 2010). In the current work, participants believed they interacted with a single individual who in some cases represented a *cooperator* and in other cases a *competitor*. This design therefore allowed us to isolate key differences between these two fundamental social contexts. First, we found that ToM regions were overall recruited more when people were more motivated to consider the other player's mental states—in this case, when the outcome of an interaction was person-dependent (active condition) versus computer-dependent (passive condition). Second, we found that most ToM regions responded similarly to cooperative and competitive interactions, suggesting that ToM is not engaged more for cooperation versus competition or vice versa as some prior research might predict. Third, despite similar recruitment of most ToM regions for cooperative and competitive trials, ToM regions, across the board, encoded differences between cooperative and competitive interactions in active but not passive conditions, even though participants were still either pitted against each other or aligned in the passive conditions. Overall, these findings provide evidence that active cooperative and competitive interactions elicit ToM to a similar extent and that ToM regions encode differences between cooperative and competitive mental states, but only when mental states influence outcome-relevant behavior during the interaction.

Prior work has shown that explicitly reasoning about the mental states of characters engaged in deception and cooperation alike recruits bilateral TPJ (Lissek et al., 2008). Our findings complement this prior work on observations of third-party interactions by showing that *acts* of cooperating and competing with another person elicit similar levels of activity in most ToM regions: the right TPJ as well as precuneus and dmPFC. The left TPJ, on the other hand, responded marginally more to competitive than cooperative trials ($p = 0.051$). Why we see this difference for lTPJ but not for other ToM regions is not entirely clear. As other researchers have proposed, the lTPJ may play a more general role in selecting among different representations, including but not limited to mental representations (Perner et al., 2006; Saxe and Young, 2013). We speculate that, in the context of the current study, lTPJ, in addition to responding to mental states, might also be associated with processing the number of unique goals in each trial: two opposing goals in competitive trials versus one shared goal in cooperative trials. We emphasize though that, for the most part, ToM regions did not respond preferentially to cooperative versus competitive trials or vice versa.

Intriguingly, our findings contrast with work showing greater medial prefrontal activity for competition or deception compared to cooperation (Decety et al., 2004; Lissek et al., 2008). One possible explanation for the apparent discrepancy between our work and these past studies is the greater mentalizing demands made by the competition and deception conditions used in prior experimental designs. For instance, participants in these studies had to check for mismatches in their own expectations and others' intentions in order to outcompete/deceive, but no such check was needed for cooperation (Decety et al., 2004; Lissek et al., 2008). We note though that everyday cooperative (and not just competitive) interactions commonly *do* require checking for mismatches between one's own and others' thoughts (e.g., I would like to help out, but would he *want* my help? She's doing something completely different from what I had asked—did she *misunderstand* me?). Thus, we speculate that certain cooperative and competitive situations may elicit response magnitude differences in ToM regions, while others may elicit similar response magnitudes. Whether situations belong to the former or the latter category may depend in particular on whether the situations contain the possibility of a misunderstanding or more generally any uncertainty about other people's actions (Jenkins and Mitchell, 2010). Indeed, behavioral uncertainty has been shown to elicit greater activity in ToM regions such as the dmPFC (Desmet et al., 2014; Dungan et al. 2016). In the present study, even though the participant was aware of the other player's goals (both at a "higher" level—to cooperate or to compete—and at a "lower" level—to match or mismatch) throughout the task, the participant could never be sure what the other player's response would be for either cooperative or competitive trials.

In addition to asking whether ToM regions are preferentially recruited during cooperative or competitive interactions, researchers can also ask about the representational content encoded in the spatial patterns of activity within these regions (Mur et al., 2008). Prior work has shown that the rTPJ, but no other ToM region, encodes the difference between accidental and intentional harms (Koster-Hale et al., 2013), though the rTPJ does not encode the difference between accidental and intentional acts within other moral domains such as purity (e.g., knowingly versus unknowingly committing incest; Chakroff et al., 2016). Meanwhile, in the present study, we found that ToM regions, across the board, encoded information separating cooperation from competition in active but not passive trials. Together, these findings suggest that ToM may be engaged differently depending on the social or moral context.

What key aspect of the interaction is being encoded in ToM regions? The current results help constrain answers to this question. We show that ToM regions do not simply encode goal-oriented differences (shared goal vs. opposing goals) or payoff-oriented differences (win together vs. only one person wins) during cooperative and competitive interactions given that these features were present for both active *and* passive conditions. Other work has shown that TPJ activity is modulated by the extent to which one perceives others' actions as affecting one's own behavior (Bhatt et al., 2010; Carter et al., 2012). The current study goes one step further in proposing that the ToM network as a whole encodes differences in how a person processes the mental states of the other player depending on whether they are in cooperation mode versus competition mode—but only when the person is motivated to consider the other player's mental states (in our case, when their behavior is thought to determine the outcome).

If people process mental states differently depending on whether they are cooperating versus competing, what might this difference reflect? In prior work, the motivation to predict others' actions versus the motivation to affiliate led to greater focus on agentive versus experiential mental states, respectively (Waytz and Young, 2014). Other work reveals that social context may lead to systematically different mental state inferences across group boundaries—people attribute different motivations for ingroup versus outgroup members for intergroup aggression (i.e., ingroup love versus outgroup hate), with significant consequences for conflict resolution (Waytz et al., 2014). Research on negotiation suggests that focus on different mental states leads to differential success as well; in a series of studies, focusing on the other person's thoughts, interests, and purposes (i.e., agency) helped people reach a deal, whereas focusing on the other person's feelings and emotions (i.e., experience) did not provide any unique advantage (Galinsky et al., 2008). It is possible that navigating cooperative and competitive interactions may primarily rely on understanding different motivational states (though this need not be a one-to-one-mapping). Our findings may be explained by sensitivity of ToM regions to these different aspects of people's mental states during cooperation versus competition.

Broadly consistent with this account is the pattern of in-scanner behavioral results indicating participants' use of different strategies for cooperative and competitive interactions, though it is important to remember that strategy choice was not related to neural classification. Specifically, participants' responses on cooperative trials followed that of a shape-self strategy (choosing the same shape they had chosen on the previous trial of the same condition) more so than a shape-other or win-stay lose-shift strategy. Notably, on cooperative trials, participants showed a tendency to choose the same shape they had chosen on the previous trial of the same condition regardless of whether they had won or lost on that previous trial; this suggests that participants were less likely to respond in anticipation of the other player's choice during cooperation. Instead, participants may have been attempting to signal their choices to the other player during cooperative trials, and therefore they may have been more tuned in to the other player's

capacity to receive their signals (i.e., capacity for experience). By contrast, during competitive trials, participants may have attempted to behave more randomly with the aim of being unpredictable. Thus, participants in competition mode may have been more sensitive to the other player's agentive mental states (i.e., capacities for planning and prediction). This account is also consistent with other work showing that people are more unwilling to be predicted during competitive interactions versus cooperative interactions (Ybarra et al., 2010); the focus of this prior work is not on the person making predictions of other people's behavior but rather on the target of prediction (i.e., the person whose behavior is being predicted by others). For instance, instead of focusing on how an interrogator can detect whether someone is lying, it's also important to focus on the person being interrogated and the signals they do or do not wish to send. While our participants may have been viewing themselves as targets of prediction in this case, future work should investigate the extent to which and the contexts in which people view themselves as predictors or targets of predictions—and how ToM is deployed for these different scenarios. Future work should also examine how the use of different behavioral strategies may relate to individual differences in reward drive and sensitivity to competitive contexts, both factors that prior work has shown to be important in determining behaviors in a competitive foraging task (Mobbs et al., 2013). Finally, future work should directly test the present speculation that ToM regions may be sensitive to different aspects of mental states during cooperation and competition and how consideration of different aspects of mental states affects cooperative and competitive behavior and neural patterns within the ToM network.

Limitations and future directions

First, we acknowledge that, although the spatial patterns of activity within ToM regions were significantly above chance, the effect sizes were modest. Our classification accuracies may appear low when compared to those found for memory research for instance (e.g., Rissman et al., 2010; Uncapher et al., 2015), but our classification accuracies are comparable to those found in other studies within the domain of social neuroscience (e.g., Chiu et al., 2011; Kaul et al., 2011; Ratner et al., 2013). Factors explaining the modest classification accuracies may be uncovered as more social neuroscientists employ multivariate analyses in their investigations.

Second, we note that passive cooperative and passive competitive trials may not seem truly cooperative or competitive. However, we point out that even though players aren't actively working together or against each other, they are placed in a cooperative or competitive stance. Namely, in passive competitive trials, players are pitted against each other: one person must lose in order for the other person to win. This is not the case for passive cooperative trials. Instead, players' successes are yoked, such that one person cannot win unless the other person does as well.

Finally, we recognize here that there are many ways in which cooperation and competition can be operationalized. Cooperation and competition can occur in many different forms and sometimes even simultaneously (e.g., cooperating with ingroup members to compete against outgroup members), all of which are outside the scope of the present study. One line of future inquiry might examine the precise ways in which cooperation and competition are operationalized, which currently varies widely by researcher and by field (Noë, 2006; Taborsky, 2007; West et al., 2007). People may cooperate with others to achieve a common goal by coordinating behavior and "acting together" (Taborsky, 2007). There is even some evidence showing coherence between signals in the superior frontal cortices of two people involved in a cooperative (but not competitive) task (Cui et al., 2012). By contrast, cooperation may occur in the form of collaboration, mismatching, or anti-coordination—in which individuals choose different actions to achieve a common goal (Abele and Stasser, 2008; Abele et al., 2014). For instance, people may work together toward an overarching goal by individually performing different tasks (e.g., one person writes and one person draws; together they make a picture book). Despite differences among these forms of cooperation, recent evidence suggests that cooperative behavior in one context is correlated with cooperative behavior in other contexts (Peysakhovich et al., 2014). The current study also provides some evidence for similarities among coordination and anti-coordination forms of cooperation. We found that most ToM regions were recruited similarly for cooperate-match trials (coordination: both players' have to match each other's shape) and cooperate-mismatch trials (anti-coordination: both players' have to mismatch—or choose the opposite shape of—the other player; $ps < 0.05$, except for precuneus: $t(14) = -2.241$, $p = 0.042$); furthermore, classification accuracy for cooperate-match and cooperate-mismatch trials was at chance level in ToM regions (all $ps > 0.05$; marginal for precuneus: $t(14) = 1.637$, $p = 0.062$). Nevertheless, whether the present results extend to different types of cooperative behavior is a worthy topic for future investigation.

Conclusion

In sum, we propose that, while people are motivated to consider others' mental states for both cooperation and competition, ToM regions encode different aspects of people's mental states during cooperation versus competition. This work thus contributes to a more detailed understanding of the neural mechanisms of ToM for cooperation and competition.

Acknowledgments

We thank members of the Boston College Morality Lab, Elizabeth Kensinger, and Scott Slotnick for their feedback and Yune-Sang Lee for his multivariate pattern analysis scripts, which we adapted and modified. This work was supported by the Alfred P. Sloan Foundation (L.Y.; grant number BR2012), the Dana Foundation (L.Y.), and National Science Foundation Graduate Research Fellowships (L.T. and J.D.; grant number 1258923).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.069.

References

Abele, S., Stasser, G., 2008. Coordination success and interpersonal perceptions: matching versus mismatching. J. Pers. Soc. Psychol. 95 (3), 576–592. http://dx.doi.org/10.1037/0022-3514.95.3.576.

Abele, S., Stasser, G., Chartier, C., 2014. Use of social knowledge in tacit coordination: social focal points. Organ. Behav. Hum. Decis. Process. 123 (1), 23–33. http://dx.doi.org/10.1016/j.obhdp.2013.10.005.

Adolphs, R., 2010. Conceptual challenges and directions for social neuroscience. Neuron 65 (6), 752–767. http://dx.doi.org/10.1016/j.neuron.2010.03.006.

Ames, D.L., Fiske, S.T., 2013. Intentional harms are worse, even when they're not. Psychol. Sci. 24 (9), 1755–1762. http://dx.doi.org/10.1177/0956797613480507.

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. Science 211 (4489), 1390–1396.

Bhatt, M.A., Lohrenz, T., Camerer, C.F., Montague, P.R., 2010. Neural signatures of strategic types in a two-person bargaining game. Proc. Natl. Acad. Sci. 107 (46), 19720–19725. http://dx.doi.org/10.1073/pnas.1009625107.

Carter, R.M., Bowling, D.L., Reeck, C., Huettel, S.A., 2012. A distinct role of the temporal–parietal junction in predicting socially guided decisions. Science 337 (6090), 109–111. http://dx.doi.org/10.1126/science.1219681.

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., Young, L., 2016. When minds matter for moral judgment: intent information is neurally encoded for harmful but not impure acts. Soc. Cogn. Affect. Neurosci. 11 (3), 476–484. http://dx.doi.org/10.1093/scan/nsv131.

Chiu, Y.-C., Esterman, M., Han, Y., Rosen, H., Yantis, S., 2011. Decoding task-based attentional modulation during face categorization. J. Cogn. Neurosci. 23 (5), 1198–1204. http://dx.doi.org/10.1162/jocn.2010.21503.

Cikara, M., Bruneau, E.G., Saxe, R.R., 2011. Us and them: intergroup failures of empathy. Curr. Dir. Psychol. Sci. 20 (3), 149–153. http://dx.doi.org/10.1177/0963721411408713.

Cui, X., Bryant, D.M., Reiss, A.L., 2012. NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. NeuroImage 59 (3), 2430–2437. http://dx.doi.org/10.1016/j.neuroimage.2011.09.003.
Cushman, F., 2008. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. Cognition 108 (2), 353–380. http://dx.doi.org/10.1016/j.cognition.2008.03.006.
Cushman, F., 2015. Deconstructing intent to reconstruct morality. Curr. Opin. Psychol. 6, 97–103. http://dx.doi.org/10.1016/j.copsyc.2015.06.003.
Decety, J., Cacioppo, S., 2012. The speed of morality: a high-density electrical neuroimaging study. J. Neurophysiol. 108 (11), 3068–3072. http://dx.doi.org/10.1152/jn.00473.2012.
Decety, J., Jackson, P.L., Sommerville, J.A., Chaminade, T., Meltzoff, A.N., 2004. The neural bases of cooperation and competition: an fMRI investigation. NeuroImage 23 (2), 744–751. http://dx.doi.org/10.1016/j.neuroimage.2004.05.025.
Desmet, C., Deschrijver, E., Brass, M., 2014. How social is error observation? The neural mechanisms underlying the observation of human and machine errors. Soc. Cogn. Affect. Neurosci. 9 (4), 427–435. http://dx.doi.org/10.1093/scan/nst002.
Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. NeuroImage 55 (2), 705–712. http://dx.doi.org/10.1016/j.neuroimage.2010.12.040.
Dungan, J., Stepanovic, M., Young, L., 2016. Theory of mind for unexpected events across contexts. Soc. Cogn. Affect. Neurosci. (in press).
Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25 (4), 1325–1335. http://dx.doi.org/10.1016/j.neuroimage.2004.12.034.
Elliott, R., Völlm, B., Drury, A., McKie, S., Richardson, P., William Deakin, J.F., 2006. Co-operation with another player in a financially rewarded guessing game activates regions implicated in theory of mind. Soc. Neurosci. 1 (3–4), 385–395. http://dx.doi.org/10.1080/17470910601041358.
Epley, N., Waytz, A., Cacioppo, J.T., 2007. On seeing human: a three-factor theory of anthropomorphism. Psychol. Rev. 114 (4), 864–886. http://dx.doi.org/10.1037/0033-295X.114.4.864.
Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive cortices. Curr. Biol. 19 (12), 1028–1033. http://dx.doi.org/10.1016/j.cub.2009.04.054.
Fiske, S.T., 1980. Attention and weight in person perception: the impact of negative and extreme behavior. J. Pers. Soc. Psychol. 38 (6), 889.
Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57 (2), 109–128.
Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994. Assessing the significance of focal activations using their spatial extent: assessing focal activations by spatial extent. Hum. Brain Mapp. 1 (3), 210–220. http://dx.doi.org/10.1002/hbm.460010306.
Galinsky, A.D., Maddux, W.W., Gilin, D., White, J.B., 2008. Why it pays to get inside the head of your opponent: the differential effects of perspective taking and empathy in negotiations. Psychol. Sci. 19 (4), 378–384. http://dx.doi.org/10.1111/j.1467-9280.2008.02096.x.
Gallagher, H., Happé, F., Brunswick, N., Fletcher, P., Frith, U., Frith, C., 2000. Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. Neuropsychologia 38, 11–21. http://dx.doi.org/10.1016/S0028-3932(99)00053-6.
Gallagher, H.L., Jack, A.I., Roepstorff, A., Frith, C.D., 2002. Imaging the intentional stance in a competitive game. NeuroImage 16 (3 Pt 1), 814–821.
Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. J. Cogn. Neurosci. 19 (11), 1803–1814. http://dx.doi.org/10.1162/jocn.2007.19.11.1803.
Gray, K., Wegner, D.M., 2009. Moral typecasting: divergent perceptions of moral agents and moral patients. J. Pers. Soc. Psychol. 96 (3), 505–520. http://dx.doi.org/10.1037/a0013748.
Gray, H.M., Gray, K., Wegner, D.M., 2007. Dimensions of mind perception. Science 315 (5812), 619. http://dx.doi.org/10.1126/science.1134475.
Hackel, L.M., Looser, C.E., Van Bavel, J.J., 2014. Group membership alters the threshold for mind perception: the role of social identity, collective identification, and intergroup threat. J. Exp. Soc. Psychol. 52, 15–23. http://dx.doi.org/10.1016/j.jesp.2013.12.001.
Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2008. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc. Natl. Acad. Sci. 105 (18), 6741–6746. http://dx.doi.org/10.1073/pnas.0711099105.
Hare, B., 2001. Can competitive paradigms increase the validity of experiments on primate social cognition? Anim. Cogn. 4 (3–4), 269–280. http://dx.doi.org/10.1007/s100710100084.
Hare, B., Tomasello, M., 2004. Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. Anim. Behav. 68 (3), 571–581. http://dx.doi.org/10.1016/j.anbehav.2003.11.011.
Harris, L.T., Fiske, S.T., 2006. Social groups that elicit disgust are differentially processed in mPFC. Soc. Cogn. Affect. Neurosci. 2 (1), 45–51. http://dx.doi.org/10.1093/scan/nsl037.
Haslam, N., 2006. Dehumanization: an integrative review. Personal. Soc. Psychol. Rev. 10 (3), 252–264. http://dx.doi.org/10.1207/s15327957pspr1003_4.
Haslam, N., Loughnan, S., 2014. Dehumanization and infrahumanization. Annu. Rev. Psychol. 65 (1), 399–423. http://dx.doi.org/10.1146/annurev-psych-010213-115045.
Imhof, L.A., Fudenberg, D., Nowak, M.A., 2007. Tit-for-tat or win-stay, lose-shift? J. Theor. Biol. 247 (3), 574–580. http://dx.doi.org/10.1016/j.jtbi.2007.03.027.
Jackson, J.W., 1993. Realistic group conflict theory: a review and evaluation of the theoretical and empirical literature. Psychol. Rec. 43 (3), 395–413.
Jenkins, A.C., Mitchell, J.P., 2010. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cereb. Cortex 20 (2), 404–410. http://dx.doi.org/10.1093/cercor/bhp109.
Kamitani, Y., Sawahata, Y., 2010. Spatial smoothing hurts localization but not information: pitfalls for brain mappers. NeuroImage 49 (3), 1949–1952. http://dx.doi.org/10.1016/j.neuroimage.2009.06.040.
Kaul, C., Rees, G., Ishai, A., 2011. The gender of face stimuli is represented in multiple regions in the human brain. Front. Hum. Neurosci. 4 http://dx.doi.org/10.3389/fnhum.2010.00238.
Kelman, H.G., 1973. Violence without moral restraint: reflections on the dehumanization of victims and victimizers. J. Soc. Issues 29 (4), 25–61. http://dx.doi.org/10.1111/j.1540-4560.1973.tb00102.x.
Koster-Hale, J., Saxe, R., Dungan, J., Young, L.L., 2013. Decoding moral judgments from neural representations of intentions. Proc. Natl. Acad. Sci. 110 (14), 5648–5653. http://dx.doi.org/10.1073/pnas.1207992110.
Kozak, M.N., Marsh, A.A., Wegner, D.M., 2006. What do I think you're doing? Action identification and mind attribution. J. Pers. Soc. Psychol. 90 (4), 543–555. http://dx.doi.org/10.1037/0022-3514.90.4.543.
Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., ... Grafman, J., 2007. Neural correlates of trust. Proc. Natl. Acad. Sci. 104 (50), 20084–20089. http://dx.doi.org/10.1073/pnas.0710103104.
Leyens, J.-P., Paladino, P.M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., Gaunt, R., 2000. The emotional side of prejudice: the attribution of secondary emotions to ingroups and outgroups. Personal. Soc. Psychol. Rev. 4 (2), 186–197. http://dx.doi.org/10.1207/S15327957PSPR0402_06.
Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., ... Brüne, M., 2008. Cooperation and deception recruit different subsets of the theory-of-mind network. PLoS One 3 (4), e2023. http://dx.doi.org/10.1371/journal.pone.0002023.
Lyons, D.E., Santos, L.R., 2006. Ecology, domain specificity, and the origins of theory of mind: Is competition the catalyst? Philos Compass 1 (5), 481–492. http://dx.doi.org/10.1111/j.1747-9991.2006.00032.x.
McAuliffe, K., Dunham, Y., 2016. Group bias in cooperative norm enforcement. Philos. Trans. R. Soc. B Biol. Sci. 371 (1686), 20150073. http://dx.doi.org/10.1098/rstb.2015.0073.
McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. Proc. Natl. Acad. Sci. 98 (20), 11832–11835. http://dx.doi.org/10.1073/pnas.211415698.
Mobbs, D., Hassabis, D., Yu, R., Chu, C., Rushworth, M., Boorman, E., Dalgleish, T., 2013. Foraging under competition: the neural basis of input-matching in humans. J. Neurosci. 33 (23), 9866–9872. http://dx.doi.org/10.1523/JNEUROSCI.2238-12.2013.
Moran, J.M., Young, L.L., Saxe, R., Lee, S.M., O'Young, D., Mavros, P.L., Gabrieli, J.D., 2011. Impaired theory of mind for moral judgment in high-functioning autism. Proc. Natl. Acad. Sci. 108 (7), 2688–2692. http://dx.doi.org/10.1073/pnas.1011734108.
Morewedge, C.K., 2009. Negativity bias in attribution of external agency. J. Exp. Psychol. Gen. 138 (4), 535–545. http://dx.doi.org/10.1037/a0016796.
Mur, M., Bandettini, P.A., Kriegeskorte, N., 2008. Revealing representational content with pattern-information fMRI–an introductory guide. Soc. Cogn. Affect. Neurosci. 4 (1), 101–109. http://dx.doi.org/10.1093/scan/nsn044.
Noë, R., 2006. Cooperation experiments: coordination through communication versus acting apart together. Anim. Behav. 71 (1), 1–18. http://dx.doi.org/10.1016/j.anbehav.2005.03.037.
Nowak, M., Sigmund, K., 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature 364 (6432), 56–58. http://dx.doi.org/10.1038/364056a0.
Opotow, S., 1990. Moral exclusion and injustice: An introduction. J. Soc. Issues 46 (1), 1–20. http://dx.doi.org/10.1111/j.1540-4560.1990.tb00268.x.
Peeters, G., Czapinski, J., 1990. Positive–negative asymmetry in evaluations: the distinction between affective and informational negativity effects. Eur. Rev. Soc. Psychol. 1 (1), 33–60.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45 (1), S199–S209. http://dx.doi.org/10.1016/j.neuroimage.2008.11.007.
Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Thinking of mental and other representations: the roles of left and right temporo-parietal junction. Soc. Neurosci. 1 (3–4), 245–258. http://dx.doi.org/10.1080/17470910600989896.
Peysakhovich, A., Nowak, M.A., Rand, D.G., 2014. Humans display a "cooperative phenotype" that is domain general and temporally stable. Nat. Commun. 5, 4939. http://dx.doi.org/10.1038/ncomms5939.
Raizada, R.D.S., Lee, Y.-S., 2013. Smoothness without smoothing: Why Gaussian Naive Bayes is not naive for multi-subject searchlight studies. PLoS One 8 (7), e69566. http://dx.doi.org/10.1371/journal.pone.0069566.
Ratner, K.G., Kaul, C., Van Bavel, J.J., 2013. Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries. Soc. Cogn. Affect. Neurosci. 8 (7), 750–755. http://dx.doi.org/10.1093/scan/nss063.
Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2004. The neural correlates of theory of mind within interpersonal interactions. NeuroImage 22 (4), 1694–1703. http://dx.doi.org/10.1016/j.neuroimage.2004.04.015.
Rissman, J., Greely, H.T., Wagner, A.D., 2010. Detecting individual memories through the neural decoding of memory states and past experience. Proc. Natl. Acad. Sci. 107 (21), 9849–9854. http://dx.doi.org/10.1073/pnas.1001028107.
Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". NeuroImage 19, 1835–1842. http://dx.doi.org/10.1016/S1053-8119(03)00230-1.
Saxe, R., Young, L., 2013. Theory of mind: how brains think about thoughts. In: Ochsner, K.N., Kosslyn, S. (Eds.), The Oxford Handbook of Cognitive Neuroscience vol. 2.
Schmelz, M., Call, J., 2016. The psychology of primate cooperation and competition: a call for realigning research agendas. Philos. Trans. R. Soc. B 371 (1686), 20150067.

Slotnick, S.D., Moo, L.R., Segal, J.B., Hart, J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Brain Res. Cogn. Brain Res. 17 (1), 75–82.
Struch, N., Schwartz, S.H., 1989. Intergroup aggression: its predictors and distinctness from in-group bias. J. Pers. Soc. Psychol. 56 (3), 364–373.
Taborsky, M., 2007. Cooperation built the Tower of Babel. Behav. Process. 76 (2), 95–99. http://dx.doi.org/10.1016/j.beproc.2007.01.013.
Tomasello, M., Vaish, A., 2013. Origins of human cooperation and morality. Annu. Rev. Psychol. 64 (1), 231–255. http://dx.doi.org/10.1146/annurev-psych-113011-143812.
Uncapher, M.R., Boyd-Meredith, J.T., Chow, T.E., Rissman, J., Wagner, A.D., 2015. Goal-directed modulation of neural memory patterns: implications for fMRI-based memory detection. J. Neurosci. 35 (22), 8531–8545. http://dx.doi.org/10.1523/JNEUROSCI.5145-14.2015.
Vaish, A., Grossmann, T., Woodward, A., 2008. Not all emotions are created equal: the negativity bias in social–emotional development. Psychol. Bull. 134 (3), 383–403. http://dx.doi.org/10.1037/0033-2909.134.3.383.
Waytz, A., Young, L., 2014. Two motivations for two dimensions of mind. J. Exp. Soc. Psychol. 55, 278–283. http://dx.doi.org/10.1016/j.jesp.2014.08.001.
Waytz, A., Morewedge, C.K., Epley, N., Monteleone, G., Gao, J.-H., Cacioppo, J.T., 2010a. Making sense by making sentient: Effectance motivation increases anthropomorphism. J. Pers. Soc. Psychol. 99 (3), 410–435. http://dx.doi.org/10.1037/a0020240.
Waytz, A., Gray, K., Epley, N., Wegner, D.M., 2010b. Causes and consequences of mind perception. Trends Cogn. Sci. 14 (8), 383–388. http://dx.doi.org/10.1016/j.tics.2010.05.006.
Waytz, A., Young, L.L., Ginges, J., 2014. Motive attribution asymmetry for love vs. hate drives intractable conflict. Proc. Natl. Acad. Sci. 111 (44), 15687–15692. http://dx.doi.org/10.1073/pnas.1414146111.
West, S.A., Griffin, A.S., Gardner, A., 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20 (2), 415–432. http://dx.doi.org/10.1111/j.1420-9101.2006.01258.x.
White, R.W., 1959. Motivation reconsidered: the concept of competence. Psychol. Rev. 66 (5), 297–333. http://dx.doi.org/10.1037/h0040934.
Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. 12 (6), 900–918. http://dx.doi.org/10.1038/jcbfm.1992.127.
Ybarra, O., Keller, M.C., Chan, E., Garcia, S.M., Sanchez-Burks, J., Morrison, K.R., Baron, A.S., 2010. Being unpredictable: friend or foe matters. Soc. Psychol. Personal. Sci. 1 (3), 259–267. http://dx.doi.org/10.1177/1948550610370214.
Young, L., Saxe, R., 2008. The neural basis of belief encoding and integration in moral judgment. NeuroImage 40 (4), 1912–1920. http://dx.doi.org/10.1016/j.neuroimage.2008.01.057.
Young, L., Saxe, R., 2009. An fMRI investigation of spontaneous mental state inference for moral judgment. J. Cogn. Neurosci. 21 (7), 1396–1405. http://dx.doi.org/10.1162/jocn.2009.21137.
Young, L., Waytz, A., 2013. Mind attribution is for morality. In: Baron-Cohen, S., Lombardo, M., Tager-Flusberg, H. (Eds.), Understanding Other Minds: Perspectives from Developmental Social Neuroscience, third ed. Oxford University Press.
Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. Proc. Natl. Acad. Sci. 104 (20), 8235–8240. http://dx.doi.org/10.1073/pnas.0701408104.
Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., 2010. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc. Natl. Acad. Sci. 107 (15), 6753–6758. http://dx.doi.org/10.1073/pnas.0914826107.