The cognitive demands on cooperation in social dilemmas: An fMRI study

Griet Emonds , Carolyn H. Declerck , Christophe Boone , Everhard J. M. Vandervliet & Paul M. Parizel

The cognitive demands on cooperation in social dilemmas: An fMRI study

Griet Emonds[1], Carolyn H. Declerck[1], Christophe Boone[1], Everhard J. M. Vandervliet[2], and Paul M. Parizel[2]

[1]Faculty of Applied Economics, University of Antwerp, Antwerp, Belgium
[2]Antwerp University Hospital (UZA) and University of Antwerp, Faculty of Medicine, Edegem, Belgium

This study uses fMRI to investigate the cognitive demands of decision-making in two types of cooperation games: a prisoner's dilemma (PD) eliciting a temptation to free-ride, leading to a dominant, self-interested response, and a stag hunt (SH) that has no dominant response but offers pay-off incentives that make mutual cooperation collectively beneficial but risky. Consequently, the PD poses greater conflict between self- and collective interest, greater demands for computational reasoning to derive the optimal solution, and greater demands for mentalizing to infer the intentions of others. Consistent with these differences between the two games, the results indicate that the PD is associated with increased activity in the anterior cingulate gyrus, prefrontal cortex, parietal lobe, and temporoparietal junction. With less conflict, the demands for computation and mentalizing are reduced in the SH, and cooperation levels increase dramatically. The differences in brain activation elicited by the different incentive structures of the PD and the SH appear to be independent of individual differences in revealed social preferences.

Understanding when and why cooperation occurs is a challenge that has occupied researchers in many disciplines. A common theme in this research is that not all cooperation is alike, and that cooperative behavior may vary substantially in its cognitive demands (Brosnan & Bshary, 2010; Stevens & Hauser, 2004). When interactions provide synergy (e.g., mutualism), mimicking behavior is encouraged and mutual cooperation occurs readily because the accruing benefits comprise a shared motive. This type of cooperative behavior presumably requires little cognition and may emerge from simple operant conditioning whereby each animal is rewarded for synchronizing its behavior with another animal (Brosnan, Salwiczek, & Bshary, 2010; Bshary, Hohner, Ait-El-Djoudi, & Fricke, 2006).

However, in dilemma situations where cooperation benefits the group but at a personal cost, mutual cooperation is more difficult to establish and maintain because both the motive of greed and the fear of betrayal favor non-cooperation. Decision-making in such cooperation dilemmas requires strategizing and involves weighing immediate costs versus potential long-term (indirect) reciprocal benefits.

Comparing the prevalence of different types of cooperation in nature reveals that mutualism (coordinating each other's cooperative decision) is wide spread, while strategic cooperation (solving a mixed-motive dilemma) remains rare, even among our closest non-human relatives (Melis & Semmann, 2010). Presumably, the latter type of cooperation has

coevolved with cognitive control and social cognition, making it possible to recognize the mixed motives and infer potential benefits while minimizing the possibility of betrayal by free-riders or untrustworthy partners (Anderson & Franks, 2001; Stevens & Hauser, 2004). Yet little is known so far regarding the neural decision-making mechanisms underlying the link between cognition and the various types of cooperation (Brosnan & Bshary, 2010). Therefore, the purpose of this study is to gain more insights into the neural correlates of decision-making that accompany these different types of cooperative behaviors in nature. Is there a qualitative difference in how the brain differentiates between a cooperative situation that yields synergy and one that presents mixed motives that could account for the ubiquity of cooperation in the former and the difficulty of cooperation in the latter? To find out, we explore whether mixed motive situations require more socio-cognitive involvement at the neural level, while mutualistic exchanges rely more on automatic processing.

The study of human cooperation is often studied with economic laboratory games that have optimal solutions derived from game theory. Mutualism is typically modeled by coordination games, while the most widely studied strategic game is the prisoner's dilemma. The latter is a dominance-solvable game with only one optimal response, which can be computed with iterative thinking. When played only once, mutual defection is the Nash equilibrium in the prisoner's dilemma. Cooperation is in fact more common in coordination games, because even in single interactions they generate synergy (Boone, Declerck, & Kiyonari, 2010). The stag hunt is a simple coordination game with a pay-off structure that superficially resembles that of the prisoner's dilemma, but without a temptation to defect. The seemingly small change in the pay-off structure that turns the prisoner's dilemma into a stag hunt aligns self- and collective interest

and generates a strong incentive to cooperate. The result is a quantum effect on behavior. We investigate with fMRI how this change in pay-off structure between a prisoner's dilemma and a stag hunt, which causes this dramatic change in behavior, is differentiated by the brain. Specifically, we test whether a stag hunt might be processed more intuitively, while solving a prisoner's dilemma requires more deliberation. Before describing the actual experiment and results, we first outline specific a priori hypotheses regarding which brain regions we expect to differ when solving a prisoner's dilemma versus a stag hunt.

HYPOTHESES

Behavior in prisoner's dilemma versus stag hunt

Both the prisoner's dilemma (from now on referred to as "PD") and stag hunt (from now on referred to as "SH") have been used extensively to study cooperation (Knez & Camerer, 2000; Kollock, 1998; Liebrand, 1983; Skyrms, 2004).

In a PD (see Figure 1a), mutual cooperation yields the most efficient collective outcome. However, the highest possible pay-off can be achieved by defecting while the other player cooperates. Knowing this, a cooperative but fearful person that anticipates a greedy other player will defect too. As a result, defect is the dominant strategy in the PD, and mutual defection is the only Nash equilibrium. In this equilibrium, neither player can improve his or her individual outcome by changing decision. The "mutual defect" solution is, however, not the most efficient way to allocate the pay-offs in a PD, and it is therefore not a Pareto optimum. The Pareto-efficient solution is the mutual

a.

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	2 / 2	3 / 0
Defection	0 / 3	1 / 1

b.

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	2 / 2	1 / 0
Defection	0 / 1	1 / 1

Figure 1.   (a) Example matrices of a prisoner's dilemma and (b) a stag hunt. Columns (rows) correspond to decision options for player A (B). The numbers in each of the four cells represent the participants' pay-off for the combination of choices made by both of them. The pay-off of player A is shown in the upper right corner of each cell, and that of player B in the lower left corner.

cooperation outcome, because then no single player can change his or her decision for a better outcome without lowering the pay-off to the other. The fact that the Nash equilibrium does not coincide with the Pareto-efficient outcome creates the strong dilemma in the PD: Players are torn between wanting the outcome that yields the most efficient distribution of pay-offs (the Pareto-optimal solution of mutual cooperation), and the awareness that each player has a personal incentive to deviate from this outcome. Because the Pareto-efficient outcome is unstable, it is difficult to predict the other's behavior in the PD. The result is what has been termed the "tragedy of the commons," (Garrett, 1968), implying that, when faced with a PD, most people will opt for the collectively deficient "defect" option.

Many game theorists have focused on ways to transform pay-off structures of mixed motive games to induce more cooperation (reviewed in Bornstein, 2003). By decreasing the pay-off from unilateral defection to or below the level of the pay-off from mutual cooperation, the temptation to free-ride in the PD is eliminated and the PD is transformed to an SH (see Figure 1b). Decision-making still occurs under uncertainty, but by removing the motive of greed, the game becomes one of coordinating on each other's decision. The SH has two Nash equilibria. Mutual cooperation is the pay-off dominant equilibrium and is also Pareto efficient. This outcome yields the largest possible pay-off for both players. Because no player can improve his outcome by moving away from this equilibrium, it becomes much easier to infer the intentions of others. Mutual defection is the risk-dominant equilibrium, which is to be preferred when uncertainty regarding the partner is high, because this outcome reduces the risk of losing. Consistent with a long tradition of research in experimental games (Knez & Camerer, 2000; Kollock, 1998; Liebrand, 1983; Skyrms, 2004), we expect more cooperation in the SH than in the PD.

Neural processing in PD versus SH

The cognitive demands to solve cooperation games are likely to involve at least two types of brain functions. First, cooperation games are dilemma situations whereby *cognitive control* is needed to evaluate the costs and benefits of alternative outcomes. The ability to detect and solve conflicts generated by a dilemma has often been attributed to the frontal cortex (Carter & van Veen, 2007; Miller & Cohen, 2001). Second, a hallmark of cooperation games is that they are social interactions with a need to predict what one's partner intends to do. The neural substrates of this mentalizing ability or "theory of mind" have been well studied and localized within the *social brain*, a neural network spanning the medial prefrontal cortex, regions lying at the junction of the temporal and parietal lobe, and the amygdala (Amodio & Frith, 2006; Brothers, 1990; Frith & Frith, 2006b). We hypothesize that the involvement of *cognitive control* and the *social brain* (*social cognition*) should differ depending on whether one is confronted with a PD (eliciting mixed motives) or an SH (eliciting synergy). This hypothesis is based on three lines of reasoning.

First, on psychological grounds, the PD and SH differ in the motives they generate. Only the PD elicits mixed motives of greed and fear whereby one has to resolve the conflict between self- and collective interest. At the neural level, conflict has consistently been associated with activation in the anterior cingulate gyrus (ACC) (reviewed in Carter & van Veen, 2007). A widely accepted function of the ACC is to monitor the occurrence of conflict or cross-talk in information processing between incompatible responses in order to evaluate the need for cognitive control. According to the conflict hypothesis, the ACC acts as a signal to engage the dorsolateral prefrontal cortex (DLPFC) to overcome the conflict and perform the task effectively (Barch, Braver, Sabb, & Noll, 2000; Carter et al., 1998; Carter & van Veen, 2007). ACC and DLPFC activation have also repeatedly been reported in fMRI studies of social interaction. For example, the ACC and DLPFC are recruited during social conflicts in an economic game when the decision to implement fairness conflicts with rationality (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Similarly, the medial prefrontal cortex touching on the ACC is activated when a person is in a competing, but not in a cooperative, stance (Decety, Jackson, Sommerville, Chaminade, & Meltzoff, 2004). In a repeated PD, the ACC is believed to monitor the conflict between collective and self-interest. ACC activation increases when a player cooperates after the partner has cooperated in the previous round. By signaling her intended continued cooperation, the partner creates a conflict for the focal player by increasing the temptation to free-ride (Rilling et al., 2002). Based on these findings, we expect that the Pareto-deficient nature of the Nash equilibrium in the PD, and the conflict created by the temptation to defect will cause more ACC and DLPFC activation than in the SH.

An additional reason why the PD is more likely to involve cognitive control than the SH is because it is a dominance-solvable game with a single solution. This means that the optimal strategy can be mathematically derived by iteratively eliminating

all dominated strategies (that are worse than other possible strategies). Such iterative thinking is not useful in the SH, where the best strategy is to imitate the strategy of the other person. Based on the results of a recent study comparing the neural correlates of solving dominance-solvable games versus pure coordination games,[1] we hypothesize that, by implementing the game-theoretic procedure of dominance-solvable games, the PD should be associated with neural structures underlying deliberation, including working-memory capacity located in the middle frontal lobe (including the DLPFC) and parietal lobe (Kuo et al., 2009). Within the latter, Kuo et al. (2009) identified increased activation in the precuneus and the inferior parietal lobule (IPL). The precuneus is often associated with agency or an "intentional self" component, mental imagery, and shifting between first and third person perspective taking (Cavanna & Trimble, 2006). It is presumed to be more active in dominance-solvable games due to the high memory requirements to keep track of mental representations that are not perceptually present (Kuo et al., 2009). A similar interpretation holds for the IPL, a structure which is also consistently associated with mentalizing and decision-making in social dilemmas (Frith & Frith, 2006a; Fukui et al., 2006), and on which we will elaborate further when discussing the need for social cognition. This combined pattern of activation in the frontal and parietal lobe has also been identified in fMRI research on the working memory and planning involved in executive goal-directed operations (Fincham, Carter, van Veen, Stenger, & Anderson, 2002; Smith & Jonides, 1998), two aspects of strategic thinking that we believe are especially necessary to solve a mixed-motive PD.

Finally, the demands on social cognition to solve a PD are expected to be higher than those required to solve a coordination dilemma. While both games require mentalizing and assessing the trustworthiness of the partner in order to predict the actual outcome of the game, the cooperative incentives of the latter should make it easier to infer the intentions of the partner. The reason we believe that intentions are more apparent in the SH is that only in this game the Pareto-efficient solution of mutual cooperation is also a Nash equilibrium. Because what is good for the self is also good for the other (and vice versa), there is less need to keep track of the perspectives of self-versus-other. The cooperative incentive applies to both players alike. In contrast, the Nash equilibrium in the

PD is not Pareto efficient, creating a social dilemma and a strong conflict between collective interest and self-interest. Knowing that one's partner also faces this conflict between a desire for a collectively efficient outcome and the temptation to defect increases the uncertainty regarding her intentions and may place greater demands on social cognition. Because alter's social preferences may be different from ours, separating self from other while contemplating a decision becomes utterly important, as should be noted by greater activity in the temporoparietal junction (TPJ), as described below.

Imaging studies of social interactions that require deciding on a course of action based on the beliefs of others have repeatedly revealed activation in the superior temporal sulcus (STS) and inferior parietal lobule (IPL), both lying near the junction of the temporal and parietal lobe (Decety & Grèzes, 2006; Frith & Frith, 2006a; Ruby & Decety, 2004; Saxe & Wexler, 2005). Furthermore, the TPJ, in conjunction with the medial prefrontal cortex, has been found to be critical for solving economic games involving trust (Krueger, Grafman, & McCabe, 2008). Because of its anatomical characteristics (heteromodal association cortex) and reciprocal connections with the prefrontal cortex, the TPJ is especially well suited for processing multisensory information as well as cognitive aspects about the self (Decety & Grèzes, 2006). The posterior STS (lying just below the TPJ) has mostly been implicated in observing and simulating actions (Frith & Frith, 2006a), and is consistently activated in social interaction when decisions are made in the knowledge that the partner is human, but not when one is playing against a pre-programmed computer (Fukui et al., 2006; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004), or when the partner cannot be held accountable for her decisions (Singer, Kiebel, Winston, Dolan, & Frith, 2004). The IPL (lying just above the TPJ) is recruited in conjunction with the prefrontal cortex when there is a need to distinguish between perspectives of self and others (Decety & Sommerville, 2003). An fMRI study shows right IPL involvement when taking a third- versus first-person perspective (Ruby & Decety, 2004), and lesions of the right IPL may lead to self–other confusion (Blakemore & Frith, 2003; Jackson & Decety, 2004).

Thus, it appears that the abilities associated with TPJ activation allow the awareness that the contents of other people's minds can be different from our own (Decety & Grèzes, 2006), and this should matter when predicting the outcome of an interdependent decision. We hypothesize that these functions, and hence TPJ activation, should be especially important when one is confronted with the conflicting motives of the PD.

---

[1]The economic games in the study by Kuo et al. (2009) did not involve cooperation and defection, and therefore the dominance-solvable games they used lacked the conflict between self-interest and collective interest that is so typical of the PD.

To summarize, because the PD generates ambiguity and conflict by its mixed motives, and because it can be solved strategically by iterative thinking, we expect that it will elicit more deliberative reasoning associated with cognitive control, and more activity in the social brain, recruiting brain regions in the frontal lobe (DLPFC), ACC, precuneus, and TPJ. Conversely, the cooperative incentives of the SH generate synergy. As cooperation becomes a shared motive, we expect that the SH removes conflict and reduces the need for strategy and working memory. In addition, coordinating each other's decisions in the SH cannot be resolved by iterative thinking, but relies instead on a "meeting of minds" that involves beliefs rather than mathematics (Bhatt & Camerer, 2005). Therefore, we hypothesize that the SH will elicit more intuitive processing, possibly activating limbic and insular regions similar to the coordination games in the study by Kuo et al. (2009).

Correlations with cooperative behavior

Synergetic incentives, such as the ones embedded in the pay-off structure of the SH, are, however, not the only determinant of cooperative behavior. Individual differences in social preferences, risk aversion, or both also have a major impact on cooperation and the outcome of social dilemmas (e.g., Camerer & Fehr, 2006; Kollock, 1998; Van Lange, 2000). These preferences are revealed in the decisions people make. Risk-averse individuals are reluctant to cooperate, regardless of the type of game. Prosocial individuals cooperate because they prefer equal outcomes. Their behavior does not depend as much on the presence of incentives, but on overcoming the fear of betrayal. Self-regarding individuals, in contrast, prefer to maximize outcomes and are more likely to adapt their cooperative behavior to the presence or absence of incentives (Boone et al., 2010). Therefore, it is also possible that people approach economic games differently due to differences in preferences, substantiating the need to disentangle the influence of incentive structures and social preferences. The current study aims to identify the brain regions that are activated in a PD versus an SH irrespective of individual differences in preferences. Therefore, to test whether the difference in brain activation between the PD and the SH is independent of social and/or risk preferences, we correlate the contrast values of PD–SH to the number of cooperative decisions made by each individual for each brain region for which we have an a priori hypothesis. Non-significant correlations with cooperative behavior increase our confidence that the observed differences between PD and SH are genuinely the

result of different underlying incentive structures and not of differences in preferences.

MATERIALS AND METHODS

Participants

Twenty-eight right-handed[2] participants (ages 19 to 33, mean age = 25, 13 men) were recruited through E-mail and Web-based announcements. Monetary incentives were emphasized. None of the participants had a history of neurological or psychiatric illness, were on any relevant medication, or suffered from claustrophobia. All participants signed an informed-consent form prior to the experiment. All these procedures were approved by the Commission of Medical Ethics at the University of Antwerp.

Analysis of the behavioral data is based on 26 participants (13 men), because the data of two participants were excluded due to a possible error in recording. The fMRI analysis is based on data of 23 participants, because the images of four participants were lost due to technical problems with the scanner, and one participant had unusual anatomical measures.

Procedures

The study was introduced to participants as an investigation of the brain areas that are involved when making economic decisions. Just prior to an experimental session, written instructions explained that each participant would be playing an investment game 56 consecutive times while under the scanner. By introducing the game as an investment game, we meant to emphasize that there were monetary earnings that could be maximized. There was no mention of different types of games. For each game, a participant would be matched with a different partner identified by a number on top of every screen. Each game required a single decision: to invest or not. The combination of the participant's decision and that of the partner would determine how many points were gained or lost for that particular game. No feedback regarding the partner's decision was given, so each decision was processed independently and could not be influenced by the outcome of the previous decision. Each point had a monetary value of 10 Eurocents. At the end of the experiment, total earnings were added to an initial 10-Euro show-up

---

[2]Candidates who indicated that they were ambivalent about their handedness were asked to fill in a handedness questionnaire (Oldfield, 1971).

fee. The average earning for the entire session was 19.31 Euros.

The instructions also explained that, for practical reasons, the game partners could not actually be present during the scanning session, but that these people had already (one by one) gone through the same procedure at an earlier time. The answers they had then given were going to be used now and matched with the answers of the current participant. In reality, participants' answers were matched to a series of random answers (50% invest),[3] as has been done in similar studies (e.g., Rilling et al., 2004; Sanfey et al., 2003). Because no outcome feedback was given during the course of the experiment, this pay-off distribution could not have affected participants' decisions.

Before starting the experiment, participants also viewed a series of photographs of their alleged future partners, without any names or numbering to preserve anonymity and to prevent participants from linking a number to a specific picture. The numbers on the screens were only meant to emphasize that every game was to be viewed as a single social interaction with a new partner.

Finally, the written instructions included two practice games with 16 questions that needed to be answered correctly before proceeding. Participants then played a practice round on a laptop to experience the decision time they would be allotted during the actual experiment.

Debriefing occurred at the conclusion of the study by contacting participants by E-mail and referring them to a website where the intent, results, and procedures of the experiment in which they had participated were fully explained.

fMRI experiment

The fMRI experiment was conducted in a blocked design with three different conditions: the PD, the SH, and a control condition where subjects viewed a number and were asked to indicate whether this number was even or odd. There were seven cycles of random PD–SH–control block alternations. Each block lasted for 30 s, during which four different matrices[4] with the same game structure (PD or SH) were shown in

_____

[3]No participants raised any doubts or questioned the procedures before, during, or after the experiment.

[4]To avoid boredom or habituation during the course of the experiment, the pay-off matrices for the games were varied. All but one of the matrices for PD and SH in Appendix 1 were retrieved from previously published studies on cooperation and were matched as best as possible to have similar pay-offs to defection.

random order. These different pay-off structures are shown in Appendix 1. Every block was separated from the previous and the next block by a resting period (30 s), during which a zero-filled matrix was shown. The total scanning time for acquiring the functional data was 21 min.

Participants under the scanner viewed the experimental stimuli in a mirror mounted on top of the head coil that projected the images from a screen positioned behind the scanner. They used a push-button to indicate their decisions.

Anatomical images were acquired with a 1.5 Tesla Siemens Sonata scanner and CP head coil (Siemens, Erlangen, Germany). A T_1-weighted MP-RAGE protocol was used (256×256 matrix, 176 1-mm sagittal slices, FOV $= 256$ mm). During the same session, functional images were acquired with $T_2{}^*$-weighted EPI (TR $= 3000$ ms, TE $= 50$ ms, 384×384 image resolution, FOV $= 192$ mm, 35 4-mm slices without gap, voxel size $= 3 \times 3 \times 4$ mm^3).

Data analysis

Image analysis was conducted with BrainVoyager QX (v 1.9.9 and 2.2.0, BrainInnovation, Maastricht, The Netherlands). Images were preprocessed by means of slice time correction, using sinc interpolation, 3D motion correction with trilinear/sinc interpolation, space domain 3D spatial smoothing with 4-mm FWHM Gaussian kernel, temporal smoothing in 3D with a high pass filter of three cycles in time course, and linear trend removal. An iso-voxel step resized the functional voxels to a $3 \times 3 \times 3$-mm configuration. For spatial normalization of the images, the standard nine-parameter landmark method of Talairach and Tournoux, implemented in BrainVoyager, was used (Talairach, Tournoux, & Musolino, 1988).

For each participant, a general linear model (GLM) was created with the percent BOLD signal change as dependent variable, and with nine regressors: PD, SH, control, and six regressors correcting for motion (translation and rotation, each in three directions). All regressors were convolved with a standard gamma model of the hemodynamic impulse–response function. Subsequently, for every participant, at every voxel in the brain, a contrast was calculated between regression coefficients of interest. To determine the effects of type of game, paired-samples t-tests were performed contrasting brain activity in PD versus SH (contrast: PD–SH).

The whole-brain statistical maps were corrected for multiple comparisons by cluster size thresholding. The initial voxel-level (uncorrected) threshold

for the contrast PD–SH was set at $p < .001$. Then the thresholded maps were submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness. After a procedure of 1000 iterations—Monte Carlo simulation, following the procedures described by Forman et al. (1995) and commonly used in other studies (e.g., Krueger et al., 2007; Suzuki, Niki, Fujisaki, & Akiyama, 2011) for estimating cluster-level false-positive rates—a mini-mum cluster-size threshold of 81 anatomical voxels, which yields a cluster-level, false-positive rate of 5%, was applied to the statistical maps. The brain regions associated with the Talairach coordinates of the peak voxel of each resulting cluster were retrieved by using Talairach Client (Lancaster et al., 1997, 2000).

In order to be able to conclude that the obtained pat-tern of activation differences between PD and SH was not due to possible differences in revealed social pref-erences, we performed additional correlation analyses. For each cluster of activation in hypothesized regions, we obtained contrast values by computing the differ-ence between the parameter estimates of PD and SH. We then calculated Pearson correlations between con-trast values and the number of cooperative decisions (totaled over PD and SH).

RESULTS

Behavior

As expected, more cooperation occurred in the SH than in the PD: mean number of cooperative decisions equals 16.15 ($\pm$9.68) or 57.69% in SH, versus 10.69 ($\pm$7.92) or 38.19% in PD, $t = 4.470$, $p < .001$. These data are similar to the results of a recent study

conducted with 322 participants, yielding 59% coop-eration in SH and 39% in PD (Boone et al., 2010).

Figure 2 shows the percent cooperative decisions in PD and SH decomposed over the different matri-ces. ANOVA tests indicate that there were significant differences in cooperation neither among the four SH matrices ($F = 1.91$, $p < .135$) nor among the four PD matrices ($F = 0.664$, $p < .576$). Paired-sampled t-tests show that cooperation in SH1, SH2, and SH3 was sig-nificantly higher compared to all four PD games (all $p < .05$). Although cooperation in SH4 also differed significantly from PD4, it did not differ compared to PD1 ($t = 1.37$, $p < .183$), PD2 ($t = 0.820$, $p < .420$), and PD3 ($t = 0.902$, $p < .376$).

Finally, Figure 3 shows a scatterplot of the num-ber of cooperative decisions in AG versus PD. This reveals a strong correlation between cooperation in PD and SH (Pearson's $r = .8$). The graph further indi-cates that there is strong between subject variation in cooperation, suggesting that participants vary in averseness (those in the left lower quadrant who con-sistently defect) and prosocial preferences (those in the right upper quadrant who consistently cooperate). This substantiates the need to control for individual dif-ferences when interpreting the results of the PD–SH contrast. However, not all data points in Figure 3 fall on the diagonal, and we note here that a few data points in the left upper quadrant may correspond to yet another individual type, namely the strategic cooperators. These are the people who cooperate in SH and defect in PD. While it is worth investigat-ing the neural correlates of strategizing in the setting of cooperation, we could not pursue this topic here due to low statistical power. Only a very small num-ber of individuals fell into this category. We refer to the study by Emonds, Declerck, Boone, Vandervliet,
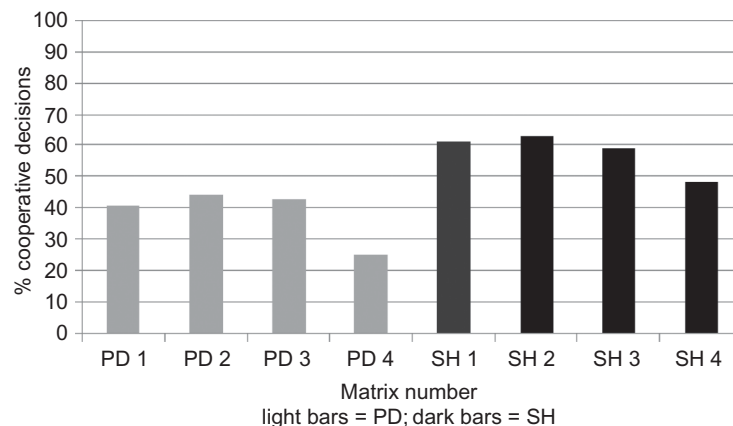

Figure 2.   Percent cooperative decisions for the different prisoner's dilemma (PD) and stag hunt (SH) pay-off matrices. PD1–PD4 or SH1–SH4 were presented in random order in each of the seven experimental blocks.
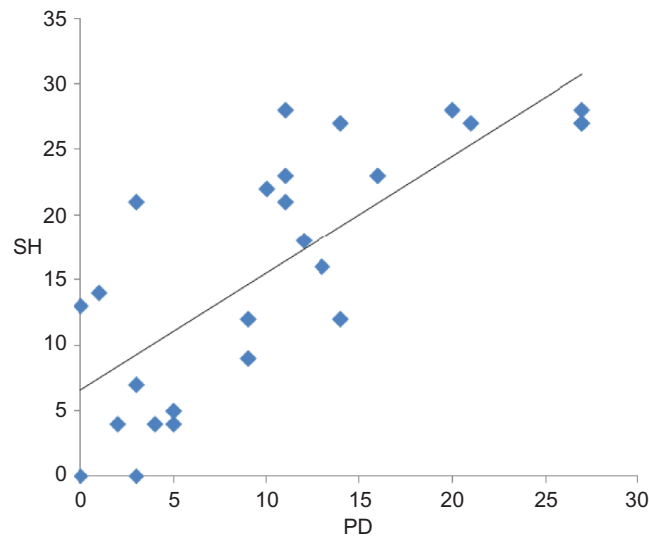
Figure 3.   Scatterplot of total number of individual cooperative decisions in the prisoner's dilemma (PD) and the stag hunt (SH), $n = 24$.

and Parizel, (2011).[5], who investigated the neural correlates of social value orientation, a personality trait which tends to correspond well with differences in strategic cooperation

fMRI results

The fMRI results for the contrast PD–SH for the whole brain are shown in Table 1. Student's t-test values for all regions yielding significant differences in activation are reported.

As hypothesized, we find greater levels of activation in the PD than the SH for the hypothesized brain regions that sustain cognitive control and deliberation: the bilateral DLPFC (BA 9, 46 within the medial frontal gyrus of the frontal lobe), the left ACC (BA 32 extending from the frontal to the limbic lobe), the precuneus (BA 7, 19 in the parietal lobe), and the bilateral IPL (BA 40 in the parietal lobe). With respect to social cognition, we find, as expected, different levels of activation between PD and SH on the border of the temporal and parietal region (TPJ), which is typically associated with mentalizing during social interaction. Specifically, the PD is associated with increased activation in the bilateral IPL (BA 40) and the right superior temporal gyrus (BA 39), which touches on the posterior STS in the temporal lobe.

---

[5]The paper we refer to is based on the same data as the current paper but discusses the analysis of neural activity with respect to individual differences in social value orientation.

Unlike expected, no brain regions were found to be more activated in the SH than in the PD. Therefore, the present data do not corroborate the hypothesis that the SH relies more on intuitive processing. In fact, Figure 4 shows that, with the exception of BA 39 (the superior temporal gyrus bordering on the posterior STS), the SH activates all hypothesized regions, but to a lesser extent than the PD. The time sequence for the percent BOLD signal change is similar in the PD and the SH, suggesting that the difference in the extent of cognitive control and social cognition needed to solve a PD and an SH is only one of degree.

Correlations with cooperative behavior

None of the correlation coefficients between contrast values of hypothesized regions and the number of cooperative decisions (see Table 1) are statistically significant at $p < .01$. This corroborates that the patterns of activation differences between PD and SH are independent of possible individual differences in behavior.

DISCUSSION

This study investigated how the brain differentiates between the pay-off structure of a PD and an SH, two economic games that are known to differ in the level of cooperative behavior they evoke. We hypothesized that the PD, where cooperation is a strategic decision, elicits more cognitive control and more social cognition

TABLE 1

Resulting *t*-values from whole brain multi-subject GLM analysis with contrast PD - SH ($t_{(22)}$, corrected $p < 0.05$, cluster size threshold = 81). Positive *t*-values refer to more activation in the PD compared to the SH. Negative *t*-values refer to more activation in the SH. Grey highlights correspond to hypothesized regions. Pearson's Correlation coefficients (*r*) relate contrast values (parameter estimates of PD – CG) and individual number of cooperative decisions. Main effect of game structure on decision making in social dilemmas (PD minus SH)

| Region	BA	Side	X	Y	Z	Size	$t_{(22)}$	r
Frontal Lobe								
Inferior Frontal Gyrus	44	R	50	10	18	114	5,253	
Middle Frontal Gyrus	9	L	−28	19	33	82	4,636	.044
		R	26	37	33	461	5,374	−.016
	46	R	44	22	21	235	5,157	−.181
	10	L	−31	43	21	470	5,991	
		R	26	49	21	337	5,538	
	6	L	−22	1	60	120	4,575	
			−31	7	48	119	4,703	
		R	29	4	54	478	5,662	
Cingulate Gyrus	32	L	−7	25	36	1343	5,805	.311
Parietal Lobe								
Inferior Parietal Lobule	40	L	−31	−50	42	340	5,377	−.099
			−40	−50	42	109	5,326	−.031
		R	38	−47	39	227	5,101	.096
Precuneus	7	L	−7	−59	36	1339	6,099	.270
			−10	−68	42	118	5,052	.374
			−22	−62	39	149	5,429	−.439
		R	26	−68	30	156	5,143	−.064
	19	L	−28	−68	30	106	4,803	−.167
Superior Parietal Lobule	7	L	−4	−65	54	84	5,053	
			−28	−71	45	323	5,497	
Temporal Lobe								
Superior Temporal Gyrus	39	R	50	−53	21	301	5,612	.055
Occipital Lobe								
Precuneus	31	R	5	−71	27	114	4,870	
Limbic Lobe								
Cingulate Gyrus	24	R	14	−5	48	140	5,533	
	31	R	5	−47	39	171	6,154	
Cingulate Gyrus	32	R	2	22	30	190	5,405	−.041
Sub-lobar Thalamus		L	−10	−32	3	131	5,399	

BA = Brodmann area; PD = Prisoner's Dilemma; SH = Stag Hunt; L = left; R = right; Size = number of 1 mm^3 voxels; Talairach coordinates and scores are taken from the peak voxel of each cluster.

than the SH, where cooperation might occur on the basis of intuition.

With respect to the contrast PD–SH, the data corroborate the hypothesis that the PD requires more cognitive control and more social cognition. Consistent with the high conflict between self-interest and collective interest in the PD, it is associated with increased activity in the bilateral DLPFC (BA 9, 46) and left ACC (BA 32). BA 9 in particular has repeatedly been linked to social tactics and strategic interactions that are so typical of a PD (Fukui et al., 2006; Krueger et al., 2007; Weissman, Perkins, & Woldorff, 2008). BA 32 involvement has also been reported in experiments in social interaction, especially when there is a conflict of interest and one is deciding between a cooperative and a defect decision (Rilling et al., 2002). Consistent with the results of Kuo et al. (2009), the PD

is also associated with more activity in regions underlying deliberative processing, including the DLPFC (BA 9, 46), the premotor area (BA 6) and frontal poles (BA 10) in the frontal lobe, the precuneus (BA 7, 19), and the bilateral IPL (BA 40) in the parietal lobe. This corroborates the hypothesis that solving a PD relies on a frontoparietal network to logically compute the single optimal solution.

The increased clusters of activation at the TPJ in the PD–SH contrast are consistent with the notion that the PD imposes greater demands on the social brain because it is a social interaction that involves predicting the decisions of the partner (Krueger et al., 2008). The increased activation cluster in BA 39 is particularly interesting, as its peak voxel ($x = 50$, $y = -53$, $z = 21$) corresponds very well to peak voxels reported in studies by Saxe and Wexler (2005)
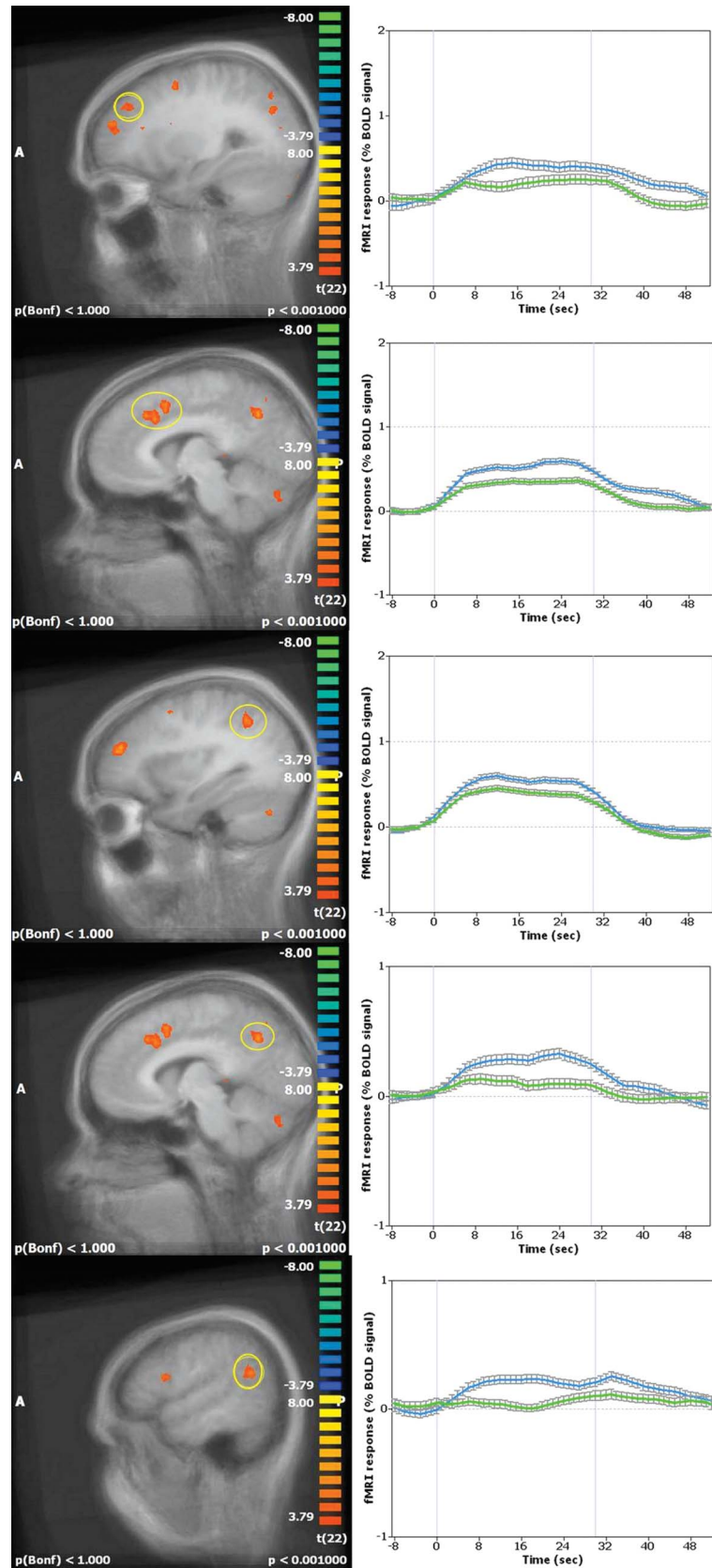
($x = 54$, $y = -54$, $z = 24$) and Bhatt, Lohrenz, Camerer, and Montague (2010) ($x = 52$, $y = -48$, $z = 20$). The study by Saxe and Wexler (2005) investigated how the social brain responds to conflicting information in a story that requires attributing a mental state to another person. The story first gave information regarding a protagonist's social background and mental state, and then concluded with a statement in which both types of information about the protagonist were either congruent or incongruent. The right TPJ (rTPJ) response (with the above coordinates) was significantly greater for incongruent information. Bhatt et al. (2010) take this finding one step further by comparing the rTPJ activation of people that use different strategies in a bargaining game. Bargainers have misaligned incentives and are able to maximize their own pay-offs by strategic deception (letting the other person believe something that conflicts with what they know––bluffing). Such strategic deception requires the ability to maintain and update another person's beliefs, relying on rTPJ activity. Importantly, in the study by Bhatt et al., the rTPJ response of strategic deceivers was sensitive to the value of the expected pay-off. The higher the value, the more incongruence there is between the knowledge of the two bargainers, implying a high degree of bluffing. The deceiver is misrepresenting information in order to instill a false belief in the other person (Bhatt et al., 2010). Understanding false belief, whether it is plainly observed (as in the study by Saxe and Wexler) or instilled (as in the study of Bhatt et al.) appears to recruit the rTPJ. The data in the current study support the notion that rTPJ is associated with recognizing the incongruence of information associated with false belief. Solving a PD requires a *strategic* analysis of the beliefs of the other player, because the goal to fulfill self-interest (leading to the Nash equilibrium) is incongruent with the goal to be collectively efficient (leading to a Pareto-optimal outcome). This misalignment of goals makes deception based on false belief possible in the PD. In contrast, in the SH, the inferred goal of the partner is congruent with that of the self, because the Nash equilibrium and the Pareto-optimal outcome are aligned. As mutual cooperation yields the largest pay-off for both, strategy and deception are not useful in the SH, relaxing the need for the rTPJ.

Finally, the increased IPL activation in the PD (where people generally cooperated less and competed more) is consistent with the finding that the competitive stance involves more IPL activity than the cooperative stance (Decety et al., 2004). These authors suggest that competing requires much less self–other merging, and thus more self–other distinction, than cooperation (see also De Cremer & Stouten, 2003). Similarly, the shared goals of the SH will make for more self–other merging and less IPL involvement, while the mixed motives of the PD will require more self–other distinction and hence more IPL involvement.

While we believe the data convincingly show that the PD elicits more neural activity in conflict resolution, strategic deliberation, and self–other distinction, they do not corroborate our hypothesis that the SH is processed more intuitively. We found no brain regions to be significantly more activated in the SH–PD contrast. This contrasts with the finding by Kuo et al. (2009) that pure coordination games involve more automatic processing. An important implication of the different results between the Kuo study and our study is that there are qualitative differences in the way different coordination games are processed. The objective in the coordination games of Kuo et al. (2009) was simply to match the number (0 to 3) chosen by another person. This choice may be influenced by features that are inferred to be salient to the other person (e.g., the belief that "many people like the number 2"), but it is unlikely to be the result of evaluating trade-offs between risks versus benefits. The SH in the present study is similar to the pure coordination game of Kuo in that the best response is also to match the choice of the other person, but differs in that the pay-off for "mutual cooperation" is not the same as the pay-off for "mutual defection." Hence evaluating the risk-averse versus the pay-off-dominant option is bound to influence decision-making in the SH and may explain why this type of coordination game still elicits cognitive processing.

Comparing the time sequence of percent Bold signal change of the PD and SH to the baseline further reveals that both games are processed in qualitatively similar ways, recruiting almost all hypothesized regions involved in cognitive control and social

---

Figure 4. *Left:* Representative brain images of group-averaged activation clusters with significant differences for the contrast prisoner's dilemma (PD) to stag hunt (SH). A = anterior; P = posterior. *Right:* Graphs plot the percent BOLD signal change averaged across subjects for the PD (blue) and the SH (green). Experimental blocks start at 0 s and end at 30 s. The baseline (0% signal change) is the mean signal in the 8-s window preceding each block. Error bars representing *SE* are depicted on each graph. *First row:* dorsolateral prefrontal cortex (BA 9), 461 voxels, sliced at voxel with peak activity ($x = 26$, $y = 37$, $z = 33$). *Second row:* anterior cingulate cortex (BA 32), 1343 voxels, sliced at voxel with peak activity ($x = -7$, $y = 25$, $z = 36$). *Third row:* inferior parietal lobe (BA 40), 340 voxels, sliced at voxel with peak activity ($x = -31$, $y = -50$, $z = 42$). *Fourth row:* Precuneus (BA 7), 1339 voxels, sliced at voxel with peak activity ($x = -7$, $y = -59$, $z = 36$). *Fifth row:* superior temporal gyrus at the TPJ (BA 39), 301 voxels, sliced at voxel with peak activity ($x = 50$, $y = -53$, $z = 21$).

cognition. We attributed these differences in the degree of activation of these regions to the different pay-off structures of the two games. Because the pay-off structure of the PD generates more conflict and less cooperative incentive, it also elicits less cooperative behavior. This finding has implications for understanding the patterns of cooperation found in nature. The ubiquity of cooperation extending all the way from bacteria to humans excludes the possibility that cooperation per se requires advanced cognition, but it is very likely that the evolutionary expansion of cognitive complexity among different species corresponds to different levels of coordination (Brosnan & Bshary, 2010). There are many documentations of coordination without cognition, such as mutualistic relations between plants and bacteria, or coordinated actions during the prey pursuit of coral reef fish or social spiders, while the cooperative hunting of vertebrates is the prime example of coordination linked to cognition. In this case, cognition becomes important because the collaborative effort requires complementarity (keeping track of each other's actions) and may also involve prey sharing. This sets the stage for possible cheating and subsequent social monitoring. The cognitive involvement that we showed to be in varying degrees associated with both types of cooperation games (SH and PD) is therefore compatible with the suggestion of Brosnan et al. (2010) that cognition is useful to cooperating animals for two reasons. First, it facilitates coordination among collaborating partners, and, second, it allows strategic decision-making concerning the best behavioral option in a given situation. This means that cognition will not always enhance cooperation, but in some cases, as in the PD, may hamper it.

A final important conclusion of the current study is that the observed differences in cognitive processing between the PD and SH are unlikely to be affected by individual differences in social or risk preferences, inasmuch as preferences are revealed in cooperative behavior. That is not to say that preferences do not affect the neural processing of decision-making, but only that the *average* differences between games that we report are probably genuine. Future research should attempt to further disentangle the differential effect of incentives (game structures) and individual differences on behavior. The challenge here is that the same behavior can result from different underlying motives. For example, to defect in a PD may be the result of risk aversion for some, while it might be a strategic judgment for others. Similarly, to cooperate in the SH may reflect the norm-abiding behavior of a prosocial person, while it is again a strategic decision serving the self-interest of a self-regarding individual. For these reasons, we believe that individuals with different preferences approach games in their own idiosyncratic way. Already there is some evidence that incentive structures matter more to self-interested individuals (Boone et al., 2010), and that they recruit more of the DLPFC when confronted with cooperation games, regardless of the type of decision they make (Emonds et al., 2011). This indicates that their heavy reliance on cognitive control is a fundamental characteristic that is not stimulus dependent. Prosocial individuals, in contrast, have been shown to elicit more amygdala activation in social interactions that involve sharing, reflecting their enhanced automatic processing mode (for intuitive decisions) and inequality aversion (Haruno & Frith, 2010).

To conclude, we believe the contribution of this study to be twofold. First, these findings shed light on the patterns of cooperation observed in nature. Humans have the computational skills to mathematically derive the best strategy in mixed motive dilemmas. This requires deliberative processing capacities relying heavily on a frontal and temporoparietal network. The DLPFC in particular developed only in full with the great expansion of the lateral regions of the brain during primate evolution (Previc, 1999), and this may have imposed constraints on the degree to which certain vertebrates can accurately assess the risk of betrayal, the benefits from future interactions, or the Pareto-efficient solution of mutual cooperation in repeated interactions (Clements & Stephens, 1995). In contrast, when interactions provide cooperators with shared benefits without the opportunity to free-ride, deliberation is less important. Such mutualistic interactions are common in nature and can in some cases be described by the logic of the SH. The data presented here show that it is not necessary to invoke a different processing mode for mutualistic interactions. The same brain functions appear to be involved in different cooperative settings, but to a much lesser extent when they involve synergy.

Second, the findings we report are also relevant to solving real-life human social dilemmas. We have shown that more cooperation can be achieved when synergetic incentives are included in the pay-off, and this also reduces the perception of conflict and the need for computation at the neural level. Furthermore, because of the shared cooperative incentives, the need for self–other differentiation is also reduced in the SH, relaxing the demands on social cognition. Future research into how social preferences affect neural networks of cooperation might reveal when cooperation is enhanced by capitalizing on extrinsic gains, when it is an internalized prosocial norms, and when it is hampered by generalized risk aversion.

REFERENCES

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition [Review]. *Nature Reviews Neuroscience*, 7(4), 268–277.

Anderson, C., & Franks, N. R. (2001). Teams in animal societies. *Behavioral Ecology*, 12(5), 534–540. doi: 10.1093/beheco/12.5.534

Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior cingulate and the monitoring of response conflict: Evidence from an fMRI study of overt verb generation. *Journal of Cognitive Neuroscience*, 12(2), 298–309.

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior*, 52(2), 424–459. doi: 10.1016/j.geb.2005.03.007

Bhatt, M., Lohrenz, T., Camerer, C., & Montague, P. R. (2010). Neural signatures of strategic types in a two-person bargaining game. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), 19780–19725. doi: 10.1073/pnas.1009625107

Blakemore, S. J., & Frith, C. (2003). Self-awareness and action [Review]. *Current Opinion in Neurobiology*, 13(2), 219–224. doi: 10.1016/s0959-4388(03)00043-6

Boone, C., Declerck, C. H., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: The moderating role of incentives and trust. *Journal of Conflict Resolution*, 54(5), 799–824. doi: 10.1177/0022002710372329

Bornstein, G. (2003). Intergroup conflict: Individual, group, and collective interests. *Personality and Social Psychology Review*, 7(2), 129–145. doi: 10.1207/S15327957PSPR0702_129-145

Brosnan, S. F., & Bshary, R. (2010). Cooperation and deception: From evolution to mechanisms. *Philosophical Transactions of the Royal Society. Series B, Biological Sciences*, 365(1553), 2593–2598. doi: 10.1098/rstb.2010.0155

Brosnan, S. F., Salwiczek, L., & Bshary, R. (2010). The interplay of cognition and cooperation. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*, 365(1553), 2699–2710. doi: 10.1098/rstb.2010.0154

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience*, 1, 27–51.

Bshary, R., Hohner, A., Ait-El-Djoudi, K., & Fricke, H. (2006). Interspecific communicative and coordinated hunting between groupers and giant moray eels in the Red Sea. *PLoS Biology*, 4(12), 2393–2398. doi: 10.1371/journal.pbio.0040431

Camerer, C. F., & Fehr, E. (2006). When does "economic man" dominate social behavior? [Review]. *Science*, 311(5757), 47–52.

Camerer, C. F., & Knez, M. (2000). Increasing cooperation in prisoner's dilemmas by establishing a precedent of efficiency in coordination. *Organizational Behavior and Human Decision Processes*, 82, 194–216. doi: 10.1006/obhd.2000.2882

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364), 747–749.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data [Review]. *Cognitive, Affective, & Behavioral Neuroscience*, 7(4), 367–379. doi: 10.3758/CABN.7.4.367

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioral correlates. *Brain*, 129, 564–583. doi: 10.1093/brain/awl004

Clements, K. C., & Stephens, D. W. (1995). Testing models of nonkin cooperation: Mutualism and the prisoner's dilemma. *Animal Behaviour*, 50, 527–535. doi: 10.1006/anbe.1995.0267

Decety, J., & Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research*, 1079, 4–14. doi: 10.1016/j.brainres.2005.12.115

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. doi: 10.1016/j.neuroimage.2004.05.025

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view [Review]. *Trends in Cognitive Sciences*, 7(12), 527–533. doi: 10.1016/j.tics.2003.10.004

De Cremer, D., & Stouten, J. (2003). When do people find cooperation most justified? The effect of trust and self-other merging in social dilemmas. *Social Justice Research*, 16(1), 41–52. doi: 10.1023/A:1022974027023

Emonds, G., Declerck, C., Boone, C., Vandervliet, E. J. M., & Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *Journal of Neuroscience, Psychology, and Economics*, 4(1), 11–24. doi: 10.1037/a0020151

Feltovich, N., Iwasaki, A., & Oda, S. H. (2011). Payoff levels, loss avoidance, and equilibrium selection in games with multiple equilibria: An experimental study. *Economic inquiry*. doi: 10.1111/j.1465-7295.2011.00406.x

Fincham, J. M., Carter, C. S., van Veen, V., Stenger, V. A., & Anderson, J. R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, 99(5), 3346–3351. doi: 10.1073/pnas.052703399

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic-resonance-imaging (fMRI) – use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33(5), 636–647. doi: 10.1002/mrm.1910330508

Frith, C. D., & Frith, U. (2006a). How we predict what other people are going to do. *Brain Research*, 1079, 36–46. doi: 10.1016/j.brainres.2005.12.126

Frith, C. D., & Frith, U. (2006b). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.

Fukui, H., Murai, T., Shinozaki, J., Aso, T., Fukuyama, H., Hayashi, T., et al. (2006). The neural basis of social tactics: An fMRI study. *NeuroImage*, 32(2), 913–920. doi: 10.1016/j.neuroimage.2006.03.039

Garrett, H. (1968). The tragedy of the commons. *Science*, 162, 1243–1248. doi: 10.1126/science.162.3859.1243

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161. doi: 10.1038/nn.2468

Jackson, P. L., & Decety, J. (2004). Motor cognition: A new paradigm to study self–other interactions [Review]. *Current Opinion in Neurobiology*, *14*(2), 259–263. doi: 10.1016/j.conb.2004.01.020

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior*, *21*, 411–427. doi: 10.1016/S1090-5138(00)00055-6

Knez, M., & Camerer, C. (2000). Increasing cooperation in prisoner's dilemmas by establishing a precedent of efficiency in coordination games. *Organizational Behavior and Human Decision Processes*, *82*(2), 194–216. doi: 10.1006/obhd.2000.2882

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation [Review]. *Annual Review of Sociology*, *24*, 183–214. doi: 10.1146/annurev.soc.24.1.183

Krueger, F., Grafman, J., & McCabe, K. (2008). Neural correlates of economic game playing [Review]. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*, *363*(1511), 3859–3874.

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20084–20089. doi: 10.1073/pnas.0710103104

Kuo, W. J., Sjostrom, T., Chen, Y. P., Wang, Y. H., & Huang, C. Y. (2009). Intuition and deliberation: Two systems for strategizing in the brain. *Science*, *324*(5926), 519–522. doi: 10.1126/science.1165598

Lancaster, J. L., Rainey, L., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., et al. (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping*, *5*, 238–242. doi: 10.1002/(SICI)1097-0193

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120–131. doi: 10.1002/1097-0193

Liebrand, W. B. G. (1983). A classification of social dilemma games. *Simulation & Gaming*, *14*(2), 123–138. doi: 10.1177/104687818301400201

Melis, A. P., & Semmann, D. (2010). How is human cooperation different? *Philosophical Transactions of the Royal Society. Series B, Biological Sciences*, *365*(1553), 2663–2674. doi: 10.1098/rstb.2010.0157

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function [Review]. *Annual Review of Neuroscience*, *24*, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Oldfield, R. C. (1971). Assessment and analysis of handedness – Edinburgh Inventory. *Neuropsychologia*, *9*(1), 97–113. doi: 10.1016/0028-3932(71)90067-4

Previc, F. H. (1999). Dopamine and the origins of human intelligence. *Brain and Cognition*, *41*(3), 299–350. doi: 10.1006/brcg.1999.1129

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395–405. doi: 10.1016/S0896-6273(02)00755-9

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport*, *15*(16), 2539–2543.

Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience*, *16*(6), 988–999. doi: 10.1162/0898929041502661

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758. doi: 10.1126/science.1082976

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–1399. doi: 10.1016/j.neuropsychologia.2005.02.013

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, *41*(4), 653–662. doi: 10.1016/S0896-6273(04)00014-5

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge, UK: Cambridge University Press.

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 12061–12068.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation [Review]. *Trends in Cognitive Sciences*, *8*(2), 60–65. doi: 10.1016/j.tics.2003.12.003

Suzuki, S., Niki, K., Fujisaki, S., & Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience*, *6*(3), 338–347. doi: 10.1093/scan/nsq042

Talairach, J., Tournoux, P., & Musolino, A. (1988). Anatomical stereotaxic studies of the frontal lobe in the management of the epilepsies. *Epilepsia*, *29*(2), 205–205.

Van Lange, P. A. M. (2000). Beyond self-interest: A set of propositions relevant to interpersonal orientations. *European Review of Social Psychology*, *11*, 297–331. doi: 10.1080/14792772043000068

Weissman, D. H., Perkins, A. S., & Woldorff, M. G. (2008). Cognitive control in social situations: A role for the dorsolateral prefrontal cortex. *NeuroImage*, *40*(2), 955–962. doi: 10.1016/j.neuroimage.2007.12.021

APPENDIX 1

PD1

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	2 \\ 2	3 \\ 0
Defection	0 \\ 3	1 \\ 1

Rilling et al., 2002

SH1

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	2 \\ 2	1 \\ 0
Defection	0 \\ 1	1 \\ 1

Kiyonari, Tanida, & Yamagishi, 2000

PD2

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	5 \\ 5	6 \\ 0
Defection	0 \\ 6	1 \\ 1

Rilling et al., 2004

SH2

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	5 \\ 5	2 \\ 0
Defection	0 \\ 2	2 \\ 2

Camerer & Knez, 2000

PD3

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	1 \\ 1	2 \\ −2
Defection	−2 \\ 2	−1 \\ −1

SH3

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	1 \\ 1	−1 \\ −5
Defection	−5 \\ −1	−1 \\ −1

Feltovich, Iwasaki, & Oda, 2008

Figure A1.   Pay-off matrices for the four prisoner's dilemma (PD) and stag hunt (SH) games presented in random order during each block in the fMRI experiment. Percent cooperation for each matrix is shown in Figure 2.

PD4

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	5 / 5	8 / 0
Defection	0 / 8	2 / 2

Camerer & Knez, 2000

SH4

Player B's choice	Player A's choice	
	Cooperation	Defection
Cooperation	7 / 7	5 / 1
Defection	1 / 5	5 / 5

Feltovivh, Iwasaki, & Oda, 2008

Figure A1. (Continued).