frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 17 December 2013
doi: 10.3389/fnins.2013.00239

Neural correlate of human reciprocity in social interactions

***Shiro Sakaiya*[1†], *Yuki Shiraito*[2,3†], *Junko Kato*[3*], *Hiroko Ide*[4], *Kensuke Okada*[5], *Kouji Takano*[6] and *Kenji Kansaku*[6]**

[1] *Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo, Tokyo, Japan*
[2] *Department of Politics, Princeton University, Princeton, NJ, USA*
[3] *Graduate School of Law and Politics, The University of Tokyo, Tokyo, Japan*
[4] *Graduate School of Humanities and Social Sciences, University of Tsukuba, Tsukuba, Japan*
[5] *Department of Psychology, Senshu University, Kawasaki, Japan*
[6] *Systems Neuroscience Section, Department of Rehabilitation for Brain Functions, Research Institute of National Rehabilitation Center for Persons with Disabilities, Tokorozawa, Japan*

Edited by:
Daeyeol Lee, Yale University School of Medicine, USA

Reviewed by:
Shunsuke Kobayashi, Fukushima Medical University, Japan
Jaeseung Jeong, Korea Advanced Institute of Science and Technology (KAIST), South Korea

****Correspondence:***
Junko Kato, Graduate School of Law and Politics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
e-mail: katoj@j.u-tokyo.ac.jp

†These authors have contributed equally to this work.

Reciprocity plays a key role maintaining cooperation in society. However, little is known about the neural process that underpins human reciprocity during social interactions. Our neuroimaging study manipulated partner identity (computer, human) and strategy (random, tit-for-tat) in repeated prisoner's dilemma games and investigated the neural correlate of reciprocal interaction with humans. Reciprocal cooperation with humans but exploitation of computers by defection was associated with activation in the left amygdala. Amygdala activation was also positively and negatively correlated with a preference change for human partners following tit-for-tat and random strategies, respectively. The correlated activation represented the intensity of positive feeling toward reciprocal and negative feeling toward non-reciprocal partners, and so reflected reciprocity in social interaction. Reciprocity in social interaction, however, might plausibly be misinterpreted and so we also examined the neural coding of insight into the reciprocity of partners. Those with and without insight revealed differential brain activation across the reward-related circuitry (i.e., the right middle dorsolateral prefrontal cortex and dorsal caudate) and theory of mind (ToM) regions [i.e., ventromedial prefrontal cortex (VMPFC) and precuneus]. Among differential activations, activation in the precuneus, which accompanied deactivation of the VMPFC, was specific to those without insight into human partners who were engaged in a tit-for-tat strategy. This asymmetric (de)activation might involve specific contributions of ToM regions to the human search for reciprocity. Consequently, the intensity of emotion attached to human reciprocity was represented in the amygdala, whereas insight into the reciprocity of others was reflected in activation across the reward-related and ToM regions. This suggests the critical role of mentalizing, which was not equated with reward expectation during social interactions.

Keywords: cooperation, reciprocity, theory of mind, reward, default mode network, amygdala, DLPFC, VMPFC

INTRODUCTION

Cooperation is a critical component of human social behavior. Compared with other social animals, humans engage in cooperative behaviors in a wide variety of social contexts (Hammerstein, 2003; Stevens and Hauser, 2004). In the prisoner's dilemma (PD) game, which has been used to examine cooperative behavior theoretically, the Nash equilibrium refers to mutual defection, which results in a lower payoff for each player than does mutual cooperation and so poses a dilemma. The situation simulated by the PD game can be replicated in society. Rational consideration is expected to lead individual members of a group, society, and international community to defect and result in a shortage of public goods and damages collective well-being in the long run (Olson, 1965). In reality, however, human cooperation is ubiquitous. Reciprocity plays a key role maintaining cooperation in society in which individuals expect to encounter others repeatedly (Axelrod and Hamilton, 1981; Henrich et al., 2003). Indeed, recent behavioral studies reported that humans are conditional cooperators who discriminately cooperate with cooperators, but not with non-cooperators (Wedekind and Milinski, 2000; Fehr and Fischbacher, 2003; Fehr and Camerer, 2007). Reciprocity, which is unique to human society (Clutton-Brock, 2009; Connor, 2010; Melis and Semmann, 2010), is intimately connected with high social cognition (Brosnan et al., 2010) and is expected to relate to the workings of the human brain (Fehr and Rockenbach, 2004; Yamasue et al., 2009). However, the brain activity that underpins human reciprocity during social interaction has not been explored fully.

To explore the neural process behind human reciprocity, we investigated the brain activation during repeated PD games in which we manipulated the experimental conditions related to the game [i.e., partner identity (computer or human) and strategy (random or tit-for-tat)] in a 2 × 2 factorial design. Published studies that manipulated the identity of partners have found more activation of theory of mind (ToM; the ability to represent the mental states of self and others) regions when games were played against humans than against computers and reported different implications of the reward circuitry between playing with humans and computers (McCabe et al., 2001; Rilling et al., 2002, 2004; Krach et al., 2008). However, the differential activation might not necessarily have been attributable ultimately to partner identity: differential activation might also be linked with differences in observed and expected behaviors of partners with distinct identities. To differentiate their effects on brain activation, we therefore manipulated both the identity and strategy of partners. Recent experiments that focused on the effect of partner strategies explored differential activation in the dorsolateral prefrontal cortex (DLPFC) when playing with human agents who were cooperative, neutral, and non-cooperative (Suzuki et al., 2011) and found activation in the superior temporal sulcus as a function of successful adaption to reciprocal/non-reciprocal strategies of computer agents (Haruno and Kawato, 2009). In distinction, we focused on a neural process that would be associated specifically with reciprocity from humans (i.e., not with computers) and cross-examined the neural processes while controlling both identity and strategy.

Whereas neuroimaging studies using games originally focused on reward-related regions [the caudate nucleus, the orbitofrontal cortex (OFC), the anterior and posterior cingulate cortex, and the DLPFC] (Lee, 2008), an increasing number of studies (McCabe et al., 2001; Rilling et al., 2002, 2008a; Fehr and Rockenbach, 2004; Lee, 2006; Rilling and Sanfey, 2011) have reported activation of brain regions implicated in ToM, including the temporal parietal junction and the precuneus (Carrington and Bailey, 2009; Van Overwalle and Baetens, 2009). The ventromedial prefrontal cortex (VMPFC) and the amygdala have been often implicated in both ToM and reward processing (Lee, 2006; Beckmann et al., 2009; Roy et al., 2012). The decision to cooperate/defect (CD) requires one to make an inference from the behavior of others in a variety of contexts and to revise and change one's own behavior based on the anticipated reactions of others and is expected to accompany activations in ToM regions. Going beyond functional localization, experimental games have recently reported the representation of subjective values relating to inference of other's behaviors and insight into opponent's strategies in ToM region (Hampton et al., 2008; Bhatt et al., 2010; Carter et al., 2012). Whereas these published experiments used games that had no possibility of cooperation, we used repeated PD games that involved the possibility of cooperation and thus reciprocity. This enabled us to explore the neural representation of insight into reciprocity of partner strategy and subjective values attached to reciprocal and non-reciprocal partners.

The neural processes underlying reciprocity in social interaction should involve both emotion and social cognition (van den Bos et al., 2009; Strobel et al., 2011). We hypothesized that the presence and absence of reciprocity in social interaction with humans involved different emotional states, which are represented by subjective values, i.e., preferences for partners, in the amygdala. The presence and absence of reciprocity, however, may be plausibly misinterpreted, particularly when inference and insight into the intent of others is required for such an interpretation. We hypothesized that those with and without insight into reciprocity would reveal differential activation across reward-related and ToM regions and that the differential activation was not a mere reflection of reward expectation. Specifically, insight into human reciprocity has great consequences for our social life. Therefore, we also hypothesized that the neural activity specific to inferring reciprocity from humans would involve regions that have been linked to mentalizing, such as the VMPFC and precuneus.

MATERIALS AND METHODS

SUBJECTS

In total, 26 healthy volunteers (14 females, 12 males; mean age ± SD, 20.5 ± 2.1) recruited from university campuses were screened to exclude those inappropriate for magnetic resonance (MR) scanning. All participants were neurologically normal and strongly right-handed according to the Edinburgh Inventory (Oldfield, 1971). Participants completed all four sessions and complied with the movement restrictions during functional magnetic resonance imaging (fMRI) sessions. Prior to scanning, participants finished a computer tutorial, and we confirmed that they fully understood the procedure. All subjects provided written informed consent for the study. This study complied with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and was approved by the Ethics Committee of the National Rehabilitation Center for Persons with Disabilities.

PROCEDURE AND DESIGN

One female and one male researcher conducted all experiments together to control for the putative influence of gender of the experimenter on the playing behavior of participants (Skotko et al., 1974).

Experiments followed a factorial design with two factors and two levels: a partner's putative identity (computer vs. human) and a partner's controlled strategy (random vs. tit-for-tat). Four sessions were conducted, and data were collected under four contextual conditions (**Figure 1**): computer-random (CR), computer-tit-for-tat (CT), human-random (HR), and human-tit-for-tat (HT). Participants engaged in each of the four sessions in random order. Each session ended after 20–23 rounds (randomly assigned) of PD games, and participants were told that each session could terminate at any point before 30 rounds were played. Each session continued for approximately 15–20 min, with a break of 3–5 min between sessions. Partners' responses were generated by a computer algorithm. A random strategy was programmed to involve cooperation (defection) with a probability of 0.5 in each round. A tit-for-tat strategy was programmed to start with cooperation and then repeat the response given by the subject in the previous round (Axelrod and Hamilton, 1981).

Participants were told that their final reward would be proportional to the total sum of their payoffs from the games. Participants were informed about the partner's identity (i.e., computer or human), but not informed about the partner's strategy (i.e., the way in which a partner was programmed to play). Participants were photographed and explained that their pictures were shown to the same-sex human partner, whereas they were shown the partner's picture: the partner was allegedly playing

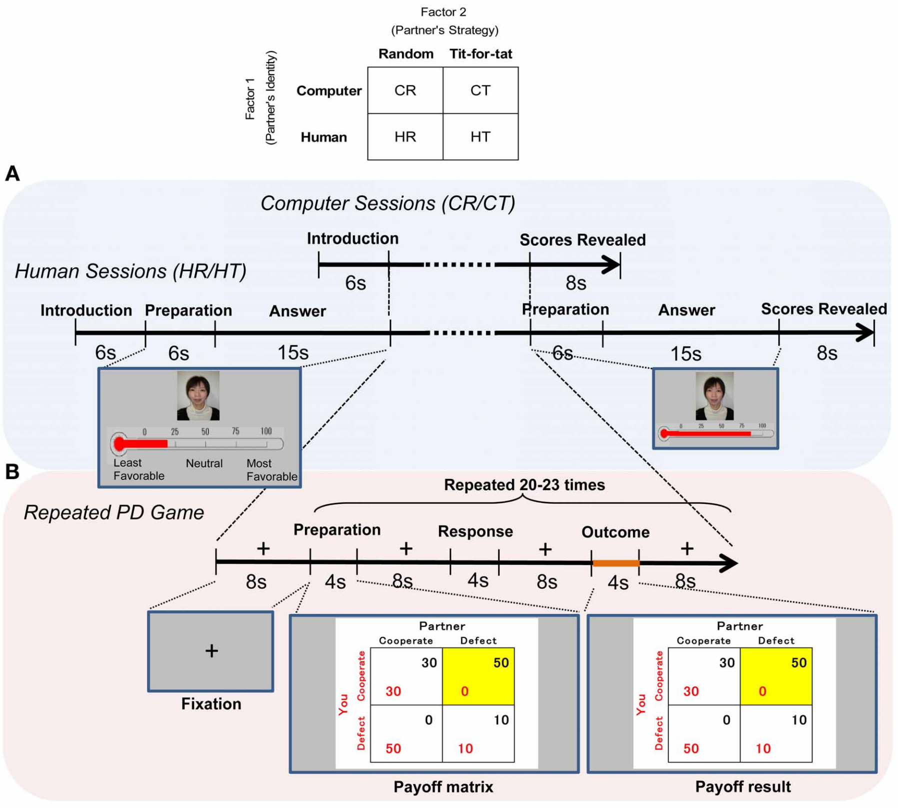

FIGURE 1 | Time course of the fMRI experiment. (A) After their partner was introduced, the subjects played 20–23 rounds of the prisoner's dilemma game, and both players' scores for the session were displayed on a screen for 8 s at the end of each of four sessions (CR/CT/HR/HT). Before and after the games with humans, the subjects viewed a picture of a same-sex partner displayed on the screen for 6 s and rated their preference on a thermometric scale for her/him for 15 s. The same picture was shown in both human (HR/HT) sessions. **(B)** During each round, the subjects viewed a payoff matrix for 4 s, pressed a key for 2 s, and viewed the outcome for 4 s. The cell of the payoff matrix chosen in the round was highlighted. Imaging analysis was focused on the outcome stage. The stages were separated by an 8-s display of a fixation cross.

in another room and was not introduced in person. Using a post-scanning questionnaire, we asked the subjects whether they had noticed any inconsistencies or differences between the actual playing of the game and the explanation prior to the experiments. All of them answered that they were playing, as had been explained, two sessions with humans and the computer, respectively; they believed that they were playing with actual humans in human sessions.

fMRI PARADIGM

At the beginning of all sessions, the subjects were informed about the identity of their partner, and this introductory stage was the same in both the computer and human sessions (see **Figure 1A** for details). Before and after playing games in each of the human (HR/HT) sessions, the subjects viewed a picture of a same-sex human partner and scaled their preferences using a thermometric scale (explained in detail in *Behavioral Data*). The same picture was shown in both human sessions. During a round of PD games in each of four sessions (**Figure 1B**), subjects chose to cooperate (C) or defect (D). Four outcomes were shown in the form of a payoff matrix: cooperate/cooperate (CC), cooperate/defect (CD), defect/cooperate (DC), and defect/defect (DD), where the first letter represents the response of participants and the second represents the response of partners. The cell of the payoff matrix chosen in the round was highlighted.

BEHAVIORAL DATA

We used two behavioral indicators for behavioral and neural analyses: changes in preference for human partners and insight into reciprocity of partner strategy. Changes in preference were aimed to quantify the participant's affection toward partners that was changed by playing games. Before and after each human (HR/HT) session, the subjects were asked to indicate their preferences. Preference was rated on a thermometric scale, on which 0° means "least favorable," 100° means "most favorable," and 50° means "neutral" (**Figure 1A**). The thermometric scale has been widely used to measure voter preferences for candidates in elections (Weisberg and Miller, 1979; Cairns et al., 2006), and was used in a published fMRI study (Kato et al., 2009). We subtracted the pre-session from the post-session rating and calculated changes in preferences. Changes in preferences were used to calculate correlations with other behavioral indicators, including the player's cooperation/defection rates and session scores (represented by the average payoff per round in each session). We also calculated their correlations with brain activation, as explained in more detail in *Neural Activity and Behavior*.

Insight into reciprocity of partner strategy was a behavioral indicator by which subjects in each session were split into two. We divided the participants into two groups according to their answer to a simple question that was focused on a critical difference between the two strategies; i.e., the presence and absence of reciprocity. After the fMRI sessions, participants who had been kept uninformed about partner strategy completed a separate questionnaire for each session to determine whether they felt that their partner had responded to their (the participants') choices. The insight group consisted of participants who answered "no" to this question for the random (CR/HR) sessions and "yes" to this question for the tit-for-tat (CT/HT) sessions. Those who answered in the opposite directions were assigned to the no-insight group. For behavioral analysis, this categorization was used as a dependent variable in logistic regression analysis that had partner's identity and strategy as independent variables. We also used it as a factor in a Three-Way analysis of variance (ANOVA) of imaging data, as explained in more details in *Imaging Analysis*.

Behavioral and psychometric analyses were conducted using the IBM SPSS Statistics (ver. 19) and Stata (ver. 11.1) software.

IMAGE ACQUISITION

The experiment was conducted in an MR scanner (Exelart, Toshiba) at the research institute. The stimulus was presented and synchronized with the MR scanner using Presentation (Neurobehavioral Systems, San Francisco, CA, USA). During sessions, gradient echo T2*-weighted echo-planar images with BOLD contrast were acquired at 1.5 T (TR/TE = 3000/40 ms, $FA = 85°$, slice thickness/gap = 6/2 mm, $FOV = 25 \times 25\,\text{cm}^2$, matrix size = 64 × 64, 18 slices). Each session consisted of 256 scans, the first three of which were discarded to allow for T1 equilibration effects. T1-weighted structural images were also acquired after the four sessions were completed.

IMAGING ANALYSIS

Imaging data were preprocessed and analyzed using SPM5 (Statistical Parametric Mapping 5; Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, UK). Realignment processing assured that the participants moved their heads less than 2 mm. The realigned images were then normalized to a Montreal Neurological Institute EPI template and smoothed with an 8-mm full-width-at-half-maximum Gaussian kernel. A high-pass filter with a cut-off period of 128 s was applied to remove low-frequency noise, and an autoregressive (order one) model was used to correct for short-range serial correlations.

We obtained individual-level contrast images during the outcome stage (4 s; **Figure 1B**), in which subjects reviewed their payoff vis-à-vis partners', since we focused on the neural correlate of reciprocity. A general linear model (GLM) was used to estimate the parameters for each experimental condition (CC, CD, DC, or DD) in four sessions (CR/CT/HR/HT). The GLM also included six additional regressors of no interest to model head movement. These regressors allowed movement effects to be discounted when looking for brain activation.

We used contrasts for each individual for the second-level analysis and estimated activation at group level. Data from 104 sessions = 26 subjects × 4 contextual conditions (CR/CT/HR/HT) were analyzed by factorial (between-subjects) ANOVA (Jackson, 2011). We performed two kinds of Three-Way ANOVA. The first model included the three factors of partner's identity, strategy, and payoff outcome [identity (computer/human) × strategy (random/tit-for-tat) × outcome (CC/CD/DC/DD)]. In the first model, we aimed to examine outcome-related activation that would vary with the identity and strategy of the partner. The second ANOVA model combined the three factors of partner's identity, strategy, and insight into reciprocity in partner strategy [identity (computer/human) × strategy (random/tit-for-tat) × insight (insight/no-insight)]. The second model enabled us to differentiate activation between those with and without insight into strategy while controlling the identity and strategy of partner. We applied a threshold of $p < 0.05$, corrected for family-wise error (FWE), for multiple comparisons across the whole brain to an activation map with the threshold of an uncorrected $p < 0.001$, combined with a cluster-size threshold of 10 voxels. Because the statistical power for detecting an interacted activation is generally low (von Eye and Schuster, 1998), we identified the appropriate regions via a *post-hoc* confirmation of statistically significant activation using contrast estimates.

We calculated contrast estimates using MarsBaR (ver. 0.43) and analyzed inferences to examine their statistical significance using IBM SPSS Statistics (ver. 19) and Stata (ver. 11.1). We used rfxplot (revision 19) (Gläscher, 2009) to generate fitted responses.

Activated regions were anatomically labeled using WFU PickAtlas (ver. 3.03, Wake Forest University School of Medicine) and SPM Anatomy Toolbox (ver. 1.8, Institute of Neuroscience and Medicine) and located visually using an anatomical atlas (Naidich et al., 2009). All x, y, and z coordinates are reported in MNI space.

NEURAL ACTIVITY AND BEHAVIOR

We used preference changes as behavioral indicators to examine the correlation with brain activation using contrast estimates. We calculated the correlation between changes in self-rated preference for human partners during each session and the individual activation in the amygdala that was associated with the entire session; i.e., estimated without a separate regressor for each round outcome. Activation patterns that were dissociated between the insight and no-insight groups were also compared for each session.

RESULTS

BEHAVIORAL ANALYSIS

Subjects cooperated more frequently in human and tit-for-tat sessions than otherwise

During the game sessions, participants chose to cooperate in an average of 58.5% of the rounds. Cooperation rates and proportions of payoff outcomes in each session were averaged over 26 participants and are summarized in Supplementary Table 1. A 2 × 2 (identity × strategy) ANOVA on cooperation rates revealed a main effect of partner identity [$F_{(1,\ 25)} = 13.98$, $p < 0.01$]. Cooperation with a human partner ($64.4 \pm 4.2\%$, mean ± SE of trials) was more frequent than that with a computer partner ($52.7 \pm 4.8\%$). We also found a main effect of strategy [$F_{(1,\ 25)} = 50.47$, $p < 0.001$]. Cooperation in tit-for-tat (CT/HT) sessions ($74.1 \pm 4.5\%$) was more likely than in random (CR/HR) sessions ($43.0 \pm 3.5\%$). The interaction between partner identity and strategy was not statistically significant.

Those with insight were found more in human and tit-for-tat sessions than otherwise

Table 1A shows the distributions of the insight and no-insight groups in each session. The logistic regression analysis (see **Table 1B** for details) showed that participants were more likely to have insight into reciprocity of partner strategy ($p < 0.05$) when playing with a human partner (i.e., HR/HT sessions) than when playing with a computer partner (i.e., CR/CT sessions). This analysis also revealed that participants tended to be more insightful ($p < 0.01$) in tit-for-tat (CT/HT) sessions than in random (CR/HR) sessions. The interaction between partner identity and strategy was not statistically significant.

Insight into reciprocity increased cooperation and payoff only in tit-for-tat sessions

Cooperation rates and proportions of payoff outcomes were significantly different between the insight and no-insight groups only in HT sessions. The cooperation rate in the no-insight group ($48.5 \pm 12.9\%$) was significantly lower [$t_{(24)} = -3.32, p < 0.01$] than that in the insight group ($86.8 \pm 5.1\%$). The rate of mutual cooperation (CC) in the no-insight group ($34.3 \pm 13.7\%$) was significantly lower [$t_{(24)} = -3.04$, $p < 0.01$] than that in the insight group ($80.6 \pm 7.3\%$), whereas the rate of mutual defection (DD) in the no-insight group ($36.4 \pm 13.1\%$) was significantly higher [$t_{(24)} = 3.47$, $p < 0.01$] than that in the insight group ($6.0 \pm 3.0\%$). In addition, session scores of the insight and no-insight groups revealed no significant differences across the four sessions [$t_{(102)} = -1.53$, $p = 0.13$] and involved no consistent difference when playing against a distinct strategy: the session scores in the no-insight group were higher [$t_{(24)} = 2.11$, $p < 0.05$] than those in the insight group in random sessions, but were significantly lower [$t_{(24)} = -3.78$, $p < 0.001$] than those in the insight group in tit-for-tat sessions. This result was consistent with the expectation that those with insight were more cooperative and rewarded more by playing games than those without insight only in tit-for-tat sessions. This also indicates that the behavioral indicator successfully distinguished those with insight from those without.

Preferences were correlated with cooperation in tit-for-tat but with payoff in random sessions

We also conducted correlation analyses on the relationship between changes in preference for partners and other behavioral indicators in human (HR/HT) sessions (see **Table 1C** for details). Preference change was significantly correlated with rate of cooperation/defection only in HT sessions: subject's cooperation rate ($r = 0.429$, $p < 0.05$) and partner's defection rate ($r = -0.436$, $p < 0.05$). We also examined the correlation between payoff outcome and preference change in each session. In HR sessions, preference change was negatively correlated with the rate of unilateral cooperation (CD; $r = -0.394$, $p < 0.05$). In HT sessions, preference change was positively correlated with the rate of mutual cooperation (CC; $r = 0.496$, $p < 0.05$) but negatively correlated with that of unilateral cooperation (CD; $r = -0.532$, $p < 0.01$) and unilateral defection (DC; $r = -0.545$, $p < 0.01$). Additionally, a preference change was significantly correlated with session scores in HR sessions only: subject's session score ($r = 0.390$, $p < 0.05$), partner's session score ($r = -0.398, p < 0.05$), and relative gain (subject's session score/partner's session score; $r = 0.417$, $p < 0.05$).

Table 1 | Results of behavioral analyses.

A. Distribution of subjects with and without insight.

Session	Insight	No-insight	Total
Computer-Random	5	21	26
Computer-TFT	15	11	26
Human-Random	12	14	26
Human-TFT	20	6	26

B. Logistic regression
(Dependent Variable: Insight = 1, No-insight = 0).

Variable	Coef.	*SE*	*Z*
Identity (Computer = 0, Human = 1)	1.59	0.63	2.51*
Strategy (Random = 0, Tit-for-tat = 1)	1.75	0.64	2.74**
Identity × Strategy	−0.70	0.88	−0.79
Constant	−1.44	0.50	−2.88**

$N = 104$, Pseudo $R^2 = 0.133$.
$^{**}p < 0.01$; $^{*}p < 0.05$.

C. Correlations with changes in preferences.

	HR session	HT session
Subject's Cooperation Rate	−0.361	0.429*
Partner's Defection Rate	−0.018	−0.436*
Mutual Cooperation (CC) Rate	−0.251	0.496*
Unilateral Cooperation (CD) Rate	−0.394*	−0.532**
Unilateral Defection (DC) Rate	0.374	−0.545**
Mutual Defection (DD) Rate	0.216	−0.263
Subject's Session Score	0.390*	0.360
Partner's Session Score	−0.398*	0.366
Relative Gain	0.417*	−0.244

Subjects with insight into their partner's strategy were those who regarded that their partners had not responded to their choices for the random (CR/HR) sessions and that their partners had responded to their choices for the tit-for-tat (CT/HT) sessions. $^{**}p < 0.01$, $^{*}p < 0.05$.

IMAGING ANALYSIS

Mutual cooperation with humans but unilateral defection with the computer was associated with activation in the amygdala

We found that outcome-related activation varied with the identity of the partner in Three-Way ANOVA [identity (computer/human) × strategy (random/tit-for-tat) × outcome (CC/CD/DC/DD)], as represented in interaction effect (identity × outcome) (Supplementary Table 2). An interaction effect was detected for a specific contrast (human$^{CC>DC}$ > computer$^{CC>DC}$) in the amygdala [(−28, −4, −14) cluster-level FWE corrected $p < 0.05$; $k = 291$] (**Figure 2A**). The amygdala was associated with mutual cooperation with humans but unilateral defection with the computer, which implied different psychological processes linked to cooperation/defection with computers/humans. A *post-hoc* examination of contrast estimates also found an interaction effect (strategy × outcome), below a threshold of multiple comparisons, in the right anterior DLPFC (Supplementary Table 2; Supplementary Figure 1).

Positive and negative subjective values were represented in the amygdala during reciprocal and non-reciprocal interactions

Individuals' different activations in the left amygdala (−28, −4, −14) in human (HR/HT) sessions were examined against subjective values, i.e., individual self-rated changes in preferences for partners. The activation was significantly negatively correlated with changes in preference in HR sessions ($r = -0.419$, $p < 0.05$) and significantly positively correlated with these changes in HT sessions ($r = 0.414$, $p < 0.05$) (**Figure 2B**).

We further examined the activation against other behavioral indicators to determine whether the relationship between the neural activity and preference changes was valid. Amygdala activation was significantly negatively correlated with relative gain (subject's session score/partner's session score) in HR sessions ($r = -0.488$, $p < 0.05$), but not HT sessions ($r = -0.356$, $p = 0.075$) (**Figure 2C**). The partial correlation between amygdala activation and relative gain was not statistically significant ($r = -0.272$, $p = 0.189$) in HR sessions when the effect of preference change was controlled. Thus, the simple correlation between amygdala activation and relative gain in HR sessions may be spurious. The neural correlate of preference changes was not significantly different when participants did and did not gain insight into the reciprocity of partners (**Figure 2B**). In HR sessions, correlations were negative in both the insight and no-insight groups ($N = 14$, $r = -0.349$, $p = 0.222$ and $N = 12$, $r = -0.350$, $p = 0.265$, respectively). In HT sessions, correlations were positive both in the insight and no-insight groups ($N = 20$, $r = 0.438$, $p = 0.053$ and $N = 6$, $r = 0.524$, $p = 0.286$). These correlations were statistically non-significant ($p > 0.05$), probably due to the limited data, but indicated that the absence of insight did not necessarily interfere with the relationship between amygdala activation and changes in preferences.

Those with and without insight into reciprocity revealed differential activation across reward circuitry and ToM regions

A Three-Way ANOVA [identity (computer/human) × strategy (random/tit-for-tat) × insight (insight/no-insight)] found differential activation between those with and without insight into strategy, while controlling the identity and strategy of partner (see **Table 2** for detailed results). Precuneus activation between the insight and no-insight groups was varied with strategy, which was represented in a significant interaction (strategy × insight) effect [BA7; (−14, −72, 32), FWE-corrected $p < 0.01$]. Insight into partner's strategy (insight > no-insight) was associated with significantly greater activity in the dorsal caudate [(6, 10, 16), FWE-corrected $p < 0.01$], VMPFC [BA32/10; (−10, 32, −4), FWE-corrected $p < 0.05$], and right middle DLPFC [BA46/9; (44, 18, 26), FWE-corrected $p < 0.05$]. Significant activation observed in the dorsal caudate and VMPFC was specific to the insight group. In contrast, significant activation in the left DLPFC [BA46; (−44, 30, 10), FWE corrected $p < 0.05$] was specific to the no-insight group. In addition to these regions in the reward circuitry and/or ToM regions, ANOVA revealed activation in the occipital cortex (Supplementary Table 3). These results demonstrate that both reward-related and ToM regions were linked to gaining insight into the partner's reciprocity during games.

Representation of insight in the DLPFC was not affected by reward expectation

Activation linked to insight might simply reflect the implications of reward processing, rather than represent insight into reciprocity in the DLPFC, which has been linked to general reasoning (Wood and Grafman, 2003; Van Overwalle, 2011), but also is a part of the reward circuitry. To explore this possibility, we examined whether differential activation was a function of payoff outcomes, which involved different material gains for subjects. Across different payoff outcomes (CC/CD/DC/DD), activation in the DLPFC was higher in the insight group than in the no-insight group for random (CR and HR) sessions (**Figure 3**). Two-sample *t*-tests of the contrast estimates for the right middle DLPFC region along outcomes showed consistently different activation patterns between the two groups in random sessions. No significant difference ($p > 0.05$) in activation, regardless of payoff outcomes, was found during tit-for-tat sessions or in the other regions implicated in random sessions.

The presence and absence of insight into human reciprocity was dissociated by precuneus activation and VMPFC deactivation

Activation in the precuneus was associated with those who did not consider that partners were responding to their prior cooperation/defection (insight group in random sessions and no-insight group in tit-for-tat sessions). **Figures 4A,B** present the interaction (strategy × insight) of precuneus activation and VMPFC activation specific to the insight group, respectively, during the human random and tit-for-tat (HR and HT) sessions. In the HT session (**Figure 4B**) specifically, significant differential activation ($p < 0.05$) in two regions was reversely associated with the presence and absence of insight. Examination of contrast estimates of differential activation revealed a significant interaction [(Insight/No-Insight) × (precuneus/VMPFC); $F_{(1,\ 51)} = 13.23$, $p < 0.001$]. The cross-over pattern of activation along the presence and absence of insight satisfies a statistical criteria for "reversed association," i.e., a "qualitative rather than quantitative difference in brain activity" to "dissociate" regions (Henson, 2005, 2006), which has been reported in recent experiments on

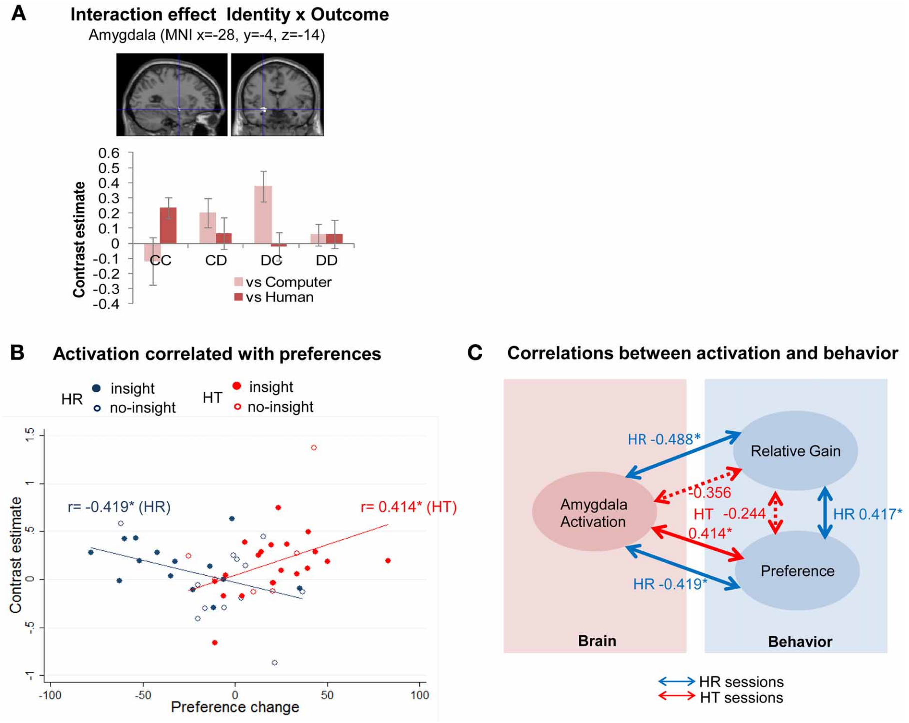

FIGURE 2 | Amygdala activation was positively and negatively correlated with preference changes for human partners following tit-for-tat and random strategies, respectively. (A) Amygdala activation was linked to mutual cooperation with humans and to unilateral defection from a computer [(−28, −4, −14), cluster-level FWE corrected $p < 0.05$]. Bar plots represent contrast estimates ± standard errors of the peak coordinates for each of the outcomes. **(B)** Participants with higher activation represented by contrast estimates at (−28, −4, −14) [cluster-level FWE corrected $p < 0.05$] showed increased preference for a partner following a tit-for-tat strategy ($r = 0.414$, $p = 0.036$), but decreased preference for one following a random strategy ($r = -0.419$, $p = 0.033$); $^*p < 0.05$. Each plot represents an individual participant. Those without insight are indicated by hollow circles. The solid blue and red lines represented the fitted regression line of all cases, i.e., those with and without insight in HR and HT, respectively. **(C)** Individual activation in the amygdala was negatively correlated with an individual's relative gain (subject's session score/partner's session score) against a human partner following a random strategy (HR) ($r = -0.488$, $p = 0.012$), whereas no statistically significant correlation was observed when playing with a human following a tit-for-tat strategy (HT) ($r = -0.356$, $p = 0.075$). Note that a partial correlation between activation of the amygdala and relative gain was not statistically significant ($r = -0.272$, $p = 0.189$) in the HR session when the effect of preference change was controlled. Dotted lines represent correlations that were not statistically significant. $^*p < 0.05$.

Table 2 | Results of Three-Way Factorial ANOVA (Identity × Strategy × Insight).

Region	Brodmann's area	Mask	MNI Coordinates			k	$F_{(1,311)}$	$t_{(311)}$
			x	y	z			
MAIN EFFECT OF INSIGHT								
Dorsal Caudate			6	10	16**	571	32.04	
VMPFC	BA32/10		−10	32	−4	229	23.03	
Right mid−DLPFC	BA46/9		44	18	26	571	22.99	
Left DLPFC	BA46		−44	30	10	96	22.02	
INTERACTION EFFECT OF STRATEGY × INSIGHT								
Precuneus	BA7		−14	−72	32**	776	28.01	
INSIGHT > NO−INSIGHT								
Dorsal Caudate		Exclusive	6	10	16**	576^{++}		5.66
VMPFC	BA32/10	Exclusive	−10	32	−4*	318^{+}		4.80
Right mid−DLPFC	BA46/9	Inclusive	44	18	26*	198		4.80
NO−INSIGHT > INSIGHT								
Left DLPFC	BA46	Exclusive	−44	30	10*	126		4.69

***FWE corrected $p < 0.01$, *FWE corrected $p < 0.05$, $^{++}$cluster-level corrected $p < 0.01$, $^{+}$cluster-level corrected $p < 0.05$.*

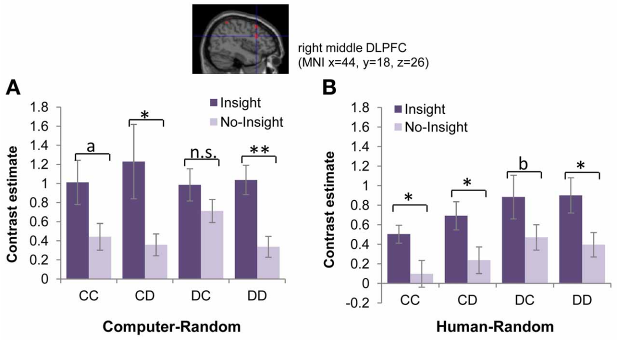

FIGURE 3 | Representation of insight into reciprocity of partner strategy in the right middle DLPFC. Significant activation in the right middle DLPFC [(44, 18, 26), FWE corrected $p < 0.05$] distinguished the insight and no-insight groups when the partner's identity (computer or human) and payoff outcome were controlled. Bar plots represent mean comparisons of activation in the right middle DLPFC between the insight and no-insight groups for each payoff outcome. Error bars represent standard errors. $^{**}p < 0.01$; $^{*}p < 0.05$; $^{a}p = 0.066$; $^{b}p = 0.067$; n.s. = not significant (two-tailed *t*-test). **(A)** Computer-random (CR) session. **(B)** Human-random (HR) session.

social cognition (Raposo et al., 2011; Zaki et al., 2012; Izuma and Adolphs, 2013). Furthermore, dissociated activation in the precuneus in the no-insight group was accompanied by significant deactivation of the VMPFC, which involved an asymmetric pattern in fitted responses of (de)activation (**Figure 4C**).

For comparison, the neural activation pattern in the computer sessions was examined (Supplementary Figure 2). Activation in the precuneus along with insight was significantly different during HT sessions, but not in CT sessions, (Supplementary Figure 2B) and so did not satisfy the criteria for "reversed association." Consequently, significant (de)activation that was dissociated across two regions was specific to those without insight into the reciprocity of human partners engaging in a tit-for-tat strategy.

DISCUSSION

RECIPROCITY IN SOCIAL INTERACTION WAS REFLECTED IN CORRELATED ACTIVATION IN THE AMYGDALA

Activation in the left amygdala was associated with mutual cooperation (CC) with humans but unilateral defection (DC) against computers (**Figure 2A**, Supplementary Table 2). Positive feelings during reciprocal cooperation with humans, but negative feelings during exploitation of computers, may involve intense affective experiences, activating the amygdala in our experiment: published experiments reported amygdala activation when trust of and reciprocity with others was undermined (Rilling et al., 2008b; Rilling and Sanfey, 2011). The left amygdala, similar to the activated region in our study, is associated with more intense social stimuli (Kramer et al., 2007) and greater emotional intensity (Phan et al., 2004), whereas the amygdala has generally been linked with both moral (Moll et al., 2005) and social (Olsson and Ochsner, 2008) cognition.

To further explore the possibility of representation of intense subjective values in the amygdala, we examined individual differential activations and found the opposite tendency in activations across sessions. Amygdala activation was positively and negatively correlated with a preference change for human partners following tit-for-tat and random strategies, respectively (**Figure 2B**). This correlation was found across the insight and no-insight groups and was thus not affected by insight into reciprocity. Examination against behavioral data also demonstrates that the correlated activation did not result from different material gains during reciprocal and non-reciprocal interactions (**Figure 2C**). Combined, these results indicate that positive feelings during reciprocal interaction but negative feelings during non-reciprocal interaction may be involved in a stronger (i.e., more intense) emotional experience than *vice versa*, which was represented in the amygdala. Behavioral analysis further confirmed this interpretation. During reciprocal interaction with human partners using a tit-for-tat strategy, mutual cooperation contributed to increasing preference vis-à-vis both unilateral cooperation and unilateral defection associated with decreased preference. Changes in preferences toward non-reciprocal human partners with a random strategy were influenced only by unilateral cooperation among behavioral outcomes, which contributed to decreasing preferences.

The amygdala has been linked to either positive or negative emotion in some studies on social decisions (Adolphs, 2002; van't Wout and Sanfey, 2008; Haruno and Frith, 2010), but has also been linked to both positive and negative social stimuli, including rewards, in others (Costafreda et al., 2008). In contrast to published works, the opposite tendency in activation in our results conveys general implications for the role of the amygdala in representing the intensity of affective values. In studies of non-social affective values such as odor (Zald and Pardo, 1997; Anderson et al., 2003) and taste (O'Doherty et al., 2001; Small et al., 2003), the intensity of values was represented in the amygdala and dissociated from valence that has been linked to the OFC. However, representation of the intensity of affective values in the amygdala has not yet been examined directly in studies on social decision, in which experimenters may not be able to apply appropriate control over the correlations and confounding effects of intensity and valence (i.e., higher intensity for negative than for positive stimuli; higher intensity and greater valence for negative stimuli) (Anderson et al., 2003). Opposite associations of valence and intensity across sessions; i.e., higher intensity for negative feelings in non-reciprocal interaction but positive feelings in reciprocal interaction, enabled our experiment to distinguish two dimensions of affective values. The representation of intensity in the amygdala, distinct from valence, in a study of social affective values was reported for the first time. The result also demonstrated that correlated activation in the amygdala reflected the presence and absence of reciprocity in social interaction.

INSIGHT INTO RECIPROCITY OF PARTNER STRATEGY WAS REPRESENTED IN REWARD CIRCUITRY AND TOM REGIONS

Differential activation was found between those with and without insight into reciprocity in partner strategy across other reward-related and ToM regions than the amygdala. The right and left DLPFC were associated with the presence and absence of

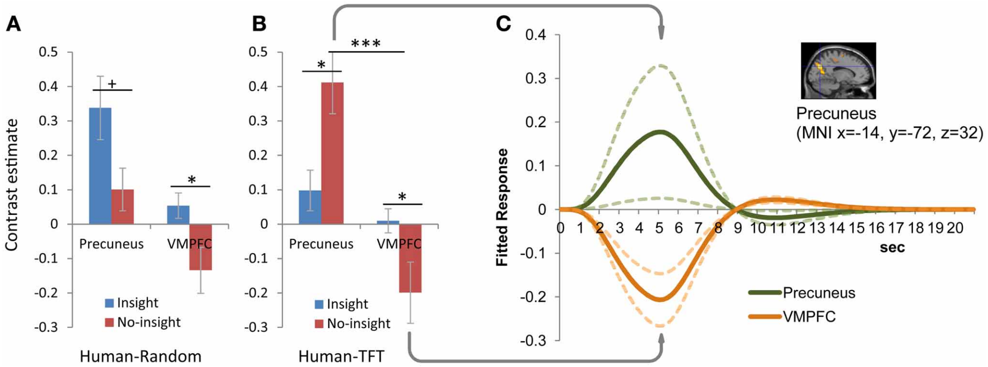

FIGURE 4 | Differential activation in the precuneus and VMPFC between those with and without insight into reciprocity of human partners. (A) Human-random (HR) session. **(B)** Human-tit-for-tat (HT) session. Insight into reciprocity interacted with partner strategy in activation in the precuneus [(−14, −72, 32), FWE corrected $p < 0.01$], but was associated specifically with activation in the VMPFC [(−10, 32, −4), FWE-corrected $p < 0.05$]. The reverse association of activation with the presence and absence of insight [(Insight/No-Insight) × (precuneus/VMPFC); $F_{(1,\ 51)} = 13.23$, $p < 0.001$] was observed for HT, but not HR sessions. Bar plots represent mean comparisons of activation in the precuneus and VMPFC between the insight and no-insight groups. Error bars represent standard errors. $^{***}p < 0.001$; $^{*}p < 0.05$; $^{+}p = 0.051$. **(C)** Asymmetric pattern of activation in the precuneus and deactivation in the VMPFC was specific to the no-insight group during HT sessions. Time-series plots (solid lines) indicate the fitted response at the coordinates. Dotted lines represent standard errors.

insight, respectively, while activation specific to insight was found in both the VMPFC and dorsal caudate (**Table 2**). A recent study reported that compliance with social norms during games activates regions in the VMPFC and the right middle DLPFC, both of which were associated with insight into reciprocity in our study (Baumgartner et al., 2011). The caudate nucleus has been implicated in trial-and-error reinforcement learning (Delgado et al., 2005), and the dorsal caudate has been associated with learning via positive and negative reinforcement and reward-seeking (O'Doherty et al., 2004; Tricomi et al., 2004; Delgado et al., 2005; Schonberg et al., 2007; Spitzer et al., 2007; Tricomi and Fiez, 2008). A region in the dorsal caudate that activated with the presence of insight in our study was implicated in a decision with uncertainty about fairness of partners in iterated trust games (Rao et al., 2013) and also with altruistic punishment of norm violation that involved no overt benefits (Strobel et al., 2011). These results have supported the possibility that activation across the reward and ToM regions might be interpreted as reflecting the involvement of subjective judgment of others rather than the expectation of reward directly from playing games.

REPRESENTATION OF INSIGHT IN THE DLPFC WAS NOT INTERFERED WITH BY REWARD EXPECTATION

Those with insight gained significantly more from playing games than those without insight in tit-for-tat sessions, but not in random sessions (see *Behavioral Analysis*); consequently, differential activations along insight might not have resulted from a simple response to material gains. To examine whether activation is a mere reflection of reward-processing, we focused on the DLPFC, which is part of the reward circuitry, but is also linked to social cognition. A recent study (De Vico Fallani et al., 2010) using the EEG hyper-scanning technique (Dumas et al., 2011) has also reported activation in this region during reciprocal interaction in the iterated PD games. In our experiment, those with insight were expected to internalize their partner's behavioral patterns, whereas those without insight more likely continued to interpret information in an effort to understand how their partners were responding. Previous reports on the right and left DLPFC are consistent with this contrasting interpretation. More specifically, the right middle DLPFC, similar to the activated region among those with insight, has been linked to judgments about interpersonal relationships (Zink et al., 2008), as well as to compliance with norms (Spitzer et al., 2007), whereas the left DLPFC, again similar to the region associated with those without insight, has been linked to integrating information (Bunge et al., 2009) and to strategic sophistication (Yoshida et al., 2010) and deception (Baumgartner et al., 2009) during games.

We further examined activation in the right DLPFC as a function of payoff outcomes and found that insight representation was not affected by reward expectation. Insight into reciprocity was represented, irrespective of payoff outcomes, in activation in random sessions (**Figures 3A,B**). As expected, we did not observe significant differential activation in tit-for-tat sessions in which the effect of insight was confounded with the one of reward. Combined, our results imply that the neural representation of insight is independent of reward expectation.

CONTRIBUTION OF TOM TO SEARCHING FOR HUMAN RECIPROCITY

Gaining insight into human reciprocity is critical for our social life. We found the brain activity that was specific to the human search for reciprocity in the VMPFC and precuneus. Although the VMPFC was thought to be generally associated with social cognition and mentalizing (Bush et al., 2000; Wood and Grafman, 2003), a region in the VMPFC that was activated among participants with insight in our study has been linked to inferences related to the intentions of others when playing games (Cooper et al., 2010). The precuneus was activated when subjects did not find partners to be responsive to their prior choices (either correctly or incorrectly, varied with the partner's actual strategies); i.e., among the insight group in random sessions and the no-insight group in tit-for-tat sessions (**Table 2**). Those without insight into reciprocity plausibly tried to discern the partner's behavioral patterns, which was consistent with the association of the precuneus with high-level integration of social cognition in general (Cavanna and Trimble, 2006). **Figures 4A,B** contrast the significant activation pattern in the precuneus and VMPFC when human partners used random and tit-for-tat strategies, respectively. The cross-over activation along insight between the two regions in a human tit-for-tat session (**Figure 4B**) involved a reversed association that represented a "qualitative" difference in the brain activity and dissociated the regions along conditions (i.e., the presence and absence of insight) (Henson, 2005, 2006). Previous experiments have found activation in both the VMPFC and precuneus during games in which participants thought that they were playing with humans rather than with computers (McCabe et al., 2001; Rilling et al., 2004; Krach et al., 2008). Distinct from this, our results indicated their further specialization in ToM. Indeed, the precuneus and VMPFC have been functionally dissociated in recent literature on the default-mode network (DMN) (Uddin et al., 2009), which is presumably activated during rest (Raichle et al., 2001) and overlaps with the ToM region (Schilbach et al., 2008; Spreng et al., 2009). The precuneus has been linked to attempts to understand the responsiveness of others, whereas the VMPFC has been linked to reasoning on true or false beliefs about reality (Sommer et al., 2007; van Buuren et al., 2010).

Those without insight into human reciprocity revealed dissociated activation, which involved asymmetric (de)activation in the precuneus and the VMPFC (**Figure 4C**). Whereas those without insight into a tit-for-tat strategy were less cooperative and rewarded than those with insight across sessions (see *Behavioral Analysis*), the significant (de)activation was observed among only those without insight into the reciprocal response of humans, but not of computers (Supplementary Figure 2). These results specifically implicate the asymmetric (de)activation in searching for human reciprocity. A recent study of a ToM-related task (i.e., episodic memory retrieval) (Sestieri et al., 2011) dissociated (de)activation of the precuneus and VMPFC, which were similar to activated regions in this study (Spreng and Grady, 2010). Taken together, the asymmetric (de)activation of the precuneus and VMPFC among those without insight into human reciprocity suggests a contribution of ToM regions specific to inferring and searching for human reciprocity.

In summary, the intensity of emotion during reciprocal and non-reciprocal interaction with humans was represented in the amygdala, whereas insight into the reciprocity of others was reflected in activation across the reward-related and ToM regions. These results indicate the critical role of mentalizing, which was not equated with reward expectation during social interactions that involved the possibility of cooperation.

AUTHOR CONTRIBUTIONS

Shiro Sakaiya, Yuki Shiraito, Junko Kato, Hiroko Ide, and Kenji Kansaku designed research; Shiro Sakaiya, Yuki Shiraito, and Kouji Takano performed research; Shiro Sakaiya, Yuki Shiraito, Junko Kato, and Kensuke Okada analyzed data; Junko Kato wrote the paper.

ACKNOWLEDGMENTS

This work was supported by the fund from Japan Society for the Promotion of Science (http://www.jsps.go.jp/english/index.html to JK 090100000119). We are grateful to Daeyeol Lee and Katsuyuki Sakai for their suggestions and comments. We also gratefully appreciate an excellent research assistance of Hiroko Nishikawa.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnins.2013.00239/abstract

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177. doi: 10.1016/S0959-4388(02)00301-X

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nat. Neurosci.* 6, 196–202. doi: 10.1038/nn1001

Axelrod, R., and Hamilton, W. D. (1981). The Evolution of Cooperation. *Science* 211, 1390–1396. doi: 10.1126/science.7466396

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., and Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron* 64, 756–770. doi: 10.1016/j.neuron.2009.11.017

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., and Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat. Neurosci.* 14, 1468–1474. doi: 10.1038/nn.2933

Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. S. (2009). Connectivity-Based Parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190. doi: 10.1523/JNEUROSCI.3328-08.2009

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

Brosnan, S. F., Salwiczek, L., and Bshary, R. (2010). The interplay of cognition and cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2699–2710. doi: 10.1098/rstb.2010.0154

Bunge, S. A., Helskog, E. H., and Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46, 338–342. doi: 10.1016/j.neuroimage.2009.01.064

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends. Cogn. Sci.* 4, 215–222. doi: 10.1016/S1364-6613(00)01483-2

Cairns, E., Kenworthy, J., Campbell, A., and Hewstone, M. (2006). The role of in-group identification, religious group membership and intergroup conflict in moderating in-group and out-group affect. *Br. J. Soc. Psychol.* 45, 701–716. doi: 10.1348/014466605X69850

Carrington, S. J., and Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum. Brain Mapp.* 30, 2313–2335. doi: 10.1002/hbm.20671

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

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

Clutton-Brock, T. (2009). Cooperation between non-kin in animal societies. *Nature* 462, 51–57. doi: 10.1038/nature08366

Connor, R. C. (2010). Cooperation beyond the dyad: on simple models and a complex society. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2687–2697. doi: 10.1098/rstb.2010.0150

Cooper, J. C., Kreps, T. A., Wiebe, T., Pirkl, T., and Knutson, B. (2010). When giving is good: ventromedial prefrontal cortex activation for others' intentions. *Neuron* 67, 511–521. doi: 10.1016/j.neuron.2010.06.030

Costafreda, S. G., Brammer, M. J., David, A. S., and Fu, C. H. Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res. Brain Res. Rev.* 58, 57–70. doi: 10.1016/j.brainresrev.2007.10.012

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575

De Vico Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., et al., (2010). Defecting or not defecting: how to "Read" human behavior during cooperative games by EEG measurements. *PLoS ONE* 5:e14187. doi: 10.1371/journal.pone.0014187

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., and George, N. (2011). From social behaviour to brain synchronization: review and perspectives in hyperscanning. *IRBM* 32, 48–53. doi: 10.1016/j.irbm.2011.01.002

Fehr, E., and Camerer, C. F. (2007). Social neuroeconornics: the neural circuitry of social preferences. *Trends. Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791. doi: 10.1038/nature02043

Fehr, E., and Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Curr. Opin. Neurobiol.* 14, 784–790. doi: 10.1016/j.conb.2004.10.007

Gläscher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics* 7, 73–82. doi: 10.1007/s12021-008-9042-x

Hammerstein, P. (2003). "Why is reciprocity so rare in social animals? A protestant appeal," in *Genetic and Cultural Evolution of Cooperation*, ed P. Hammerstein (Cambridge: MIT Press), 83–94.

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.0711099105

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

Haruno, M., and Kawato, M. (2009). Activity in the superior temporal sulcus highlights learning competence in an interaction game. *J. Neurosci.* 29, 4542–4547. doi: 10.1523/JNEUROSCI.2707-08.2009

Henrich, J., Bowles, S., Boyd, R., Hopfensitz, A., Richardson, P., Sigmund, K., et al., (2003). "The cultural and genetic evolution of human cooperation," in *Generic and Cultural Evolution of Cooperation*, ed P. Hammerstein (Cambridge: MIT Press), 445–468.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Q. J. Exp. Psychol. A* 58, 193–233. doi: 10.1080/027249804430 00502

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends. Cogn. Sci.* 10, 64–69. doi: 10.1016/j.tics.2005.12.005

Izuma, K., and Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron* 78, 563–573. doi: 10.1016/j.neuron.2013.03.023

Jackson, S. L. (2011). *Research Methods and Statistics: a Critical Thinking Approach, 4th Edn.* Belmont: Wadsworth Publishing.

Kato, J., Ide, H., Kabashima, I., Kadota, H., Takano, K., and Kansaku, K. (2009). Neural correlates of attitude change following positive and negative advertisements. *Front. Behav. Neurosci.* 3:6. doi:10.3389/neuro.3308.3006.2009

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3:e2597. doi: 10.1371/journal.pone.0002597

Kramer, U. M., Jansma, H., Tempelmann, C., and Munte, T. F. (2007). Tit-for-tat: the neural basis of reactive aggression. *Neuroimage* 38, 203–211. doi: 10.1016/j.neuroimage.2007.07.029

Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409. doi: 10.1038/nn2065

Lee, D. Y. (2006). Neural basis of quasi-rational decision making. *Curr. Opin. Neurobiol.* 16, 191–198. doi: 10.1016/j.conb.2006.02.001

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11832–11835. doi: 10.1073/pnas.211415698

Melis, A. P., and Semmann, D. (2010). How is human cooperation different? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2663–2674. doi: 10.1098/rstb.2010.0157

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., and Grafman, J. (2005). The neural basis of human moral cognition. *Nat. Rev. Neurosci.* 6, 799–809. doi: 10.1038/nrn1768

Naidich, T. P., Duvernoy, H. M., and Benner, T. (2009). *Duvernoy's Atlas of the Human Brain Stem and Cerebellum: High-Field MRI: Surface Anatomy, Internal Structure, Vascularization and 3D Sectional Anatomy*. Wien: Springer.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454. doi: 10.1126/science.1094285

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., and McGlone, F. (2001). Representation of pleasant and aversive taste in the human brain. *J. Neurophysiol.* 85, 1315–1321.

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

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*, Vol. v. Cambridge, MA: Harvard University Press.

Olsson, A., and Ochsner, K. N. (2008). The role of social cognition in emotion. *Trends. Cogn. Sci.* 12, 65–71. doi: 10.1016/j.tics.2007.11.010

Phan, K. L., Taylor, S. F., Welsh, R. C., Ho, S. H., Britton, J. C., and Liberzon, I. (2004). Neural correlates of individual ratings of emotional salience: a trial-related fMRI study. *Neuroimage* 21, 768–780. doi: 10.1016/j.neuroimage.2003.09.072

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676

Rao, H., Wardle, M. C., Fitzgerald, D. A., Angstadt, M., Sripada, C. S., McCabe, K., et al. (2013). The caudate signals bad reputation during trust decisions. *PLoS ONE* 8:e68884. doi: 10.1371/journal.pone.0068884

Raposo, A., Vicens, L., Clithero, J. A., Dobbins, I. G., and Huettel, S. A. (2011). Contributions of frontopolar cortex to judgments about self, others and relations. *Soc. Cogn. Affect. Neurosci.* 6, 260–269. doi: 10.1093/scan/nsq033

Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., et al. (2008a). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266. doi: 10.1016/j.neuropsychologia.2007.11.033

Rilling, J. K., King-Casas, B., and Sanfey, A. G. (2008b). The neurobiology of social decision-making. *Curr. Opin. Neurobiol.* 18, 159–165. doi: 10.1016/j.conb.2008.06.003

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

Rilling, J. K., and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48. doi: 10.1146/annurev.psych.121208.131647

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703. doi: 10.1016/j.neuroimage.2004.04.015

Roy, M., Shohamy, D., and Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends. Cogn. Sci.* 16, 147–156. doi: 10.1016/j.tics.2012.01.005

Schilbach, L., Eickhoff, S. B., Rska-Jagiela, A. R., Fink, G. R., and Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Conscious. Cogn.* 17, 457–467. doi: 10.1016/j.concog.2008.03.013

Schonberg, T., Daw, N. D., Joel, D., and O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J. Neurosci.* 27, 12860–12867. doi: 10.1523/JNEUROSCI.2496-07.2007

Sestieri, C., Corbetta, M., Romani, G. L., and Shulman, G. L. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: functional and topographic analyses. *J. Neurosci.* 31, 4407–4420. doi: 10.1523/JNEUROSCI.3335-10.2011

Skotko, V., Langmeyer, D., and Lundgren, D. (1974). Sex-differences as artifact in prisoner's dilemma game. *J. Conflict Resolut.* 18, 707–713. doi: 10.1177/002200277401800411

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39, 701–711. doi: 10.1016/S0896-6273(03)00467-7

Sommer, M., Dohnel, K., Sodian, B., Meinhardt, J., Thoermer, C., and Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage* 35, 1378–1384. doi: 10.1016/j.neuroimage.2007.01.042

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Spreng, R. N., and Grady, L. C. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *J. Cogn. Neurosci.* 22, 1112–1123. doi: 10.1162/jocn.2009.21282

Spreng, R. N., Mar, R. A., and Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* 21, 489–510. doi: 10.1162/jocn.2008.21029

Stevens, J. R., and Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends. Cogn. Sci.* 8, 60–65. doi: 10.1016/j.tics.2003.12.003

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., et al. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage* 54, 671–680. doi: 10.1016/j.neuroimage.2010.07.051

Suzuki, S., Niki, K., Fujisaki, S., and Akiyama, E. (2011). Neural basis of conditional cooperation. *Soc. Cogn. Affect. Neurosci.* 6, 338–347. doi: 10.1093/scan/nsq042

Tricomi, E., and Fiez, J. A. (2008). Feedback signals in the caudate reflect goal achievement on a declarative memory task. *Neuroimage* 41, 1154–1167. doi: 10.1016/j.neuroimage.2008.02.066

Tricomi, E. M., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292. doi: 10.1016/S0896-6273(03)00848-1

Uddin, L. Q., Kelly, A. M. C., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2009). Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum. Brain Mapp.* 30, 625–637. doi: 10.1002/hbm.20531

van Buuren, M., Gladwin, T. E., Zandbelt, B. B., Kahn, R. S., and Vink, M. (2010). Reduced functional coupling in the default-mode network during self-referential processing. *Hum. Brain Mapp.* 31, 1117–1127. doi: 10.1002/hbm.20920

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., and Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the trust game. *Soc. Cogn. Affect. Neurosci.* 4, 294–304. doi: 10.1093/scan/nsp009

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage* 54, 1589–1599. doi: 10.1016/j.neuroimage.2010.09.043

Van Overwalle, F., and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584. doi: 10.1016/j.neuroimage.2009.06.009

von Eye, A., and Schuster, C. (1998). *Regression Analysis for Social Sciences, 1st Edn*. Waltham, MA: Academic Press.

Wedekind, C., and Milinski, M. (2000). Cooperation through image scoring in humans. *Science* 288, 850–852. doi: 10.1126/science.288.5467.850

Weisberg, H. F., and Miller, A. H. (1979). *Evaluation of the feeling thermometer: a report to the national election study board based on data from the 1979 pilot survey*. Ann Arbor, MI: Center for Political Studies, Institute for Social Research.

Wood, J. N., and Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nat. Rev. Neurosci.* 4, 139–147. doi: 10.1038/nrn1033

van't Wout, M., and Sanfey, A. G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108, 796–803. doi: 10.1016/j.cognition.2008.07.002

Yamasue, H., Kuwabara, H., Kawakubo, Y., and Kasai, K. (2009). Oxytocin, sexually dimorphic features of the social brain, and autism. *Psychiatry Clin. Neurosci.* 63, 129–140. doi: 10.1111/j.1440-1819.2009.01944.x

Yoshida, W., Seymour, B., Friston, K. J., and Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 30, 10744–10751. doi: 10.1523/JNEUROSCI.5895-09.2010

Zaki, J., Davis, J. I., and Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *Neuroimage* 62, 493–499. doi: 10.1016/j.neuroimage.2012.05.012

Zald, D. H., and Pardo, J. V. (1997). Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4119–4124. doi: 10.1073/pnas.94.8.4119

Zink, C. F., Tong, Y. X., Chen, Q., Bassett, D. S., Stein, J. L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* 58, 273–283. doi: 10.1016/j.neuron.2008.01.025

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 September 2013; accepted: 27 November 2013; published online: 17 December 2013.

Citation: Sakaiya S, Shiraito Y, Kato J, Ide H, Okada K, Takano K and Kansaku K (2013) Neural correlate of human reciprocity in social interactions. Front. Neurosci. 7:239. doi: 10.3389/fnins.2013.00239

This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience.