Experimental Social Sciences
Toward Experimentally-based New Social Sciences for the 21st Century

ExpSS21

Working Paper

ExpSS-E-10

The Cognitive Processing of Altruistic and Spiteful Behavior: An fMRI Study

Tsuyoshi Nihonsugi
Institute of Social and Economic Research, Osaka University,
JSPS Research Fellow/ISER, Osaka University

Takafumi Yamakawa
Institute of Social and Economic Research, Osaka University

Jun Shinozaki
Institute of Social and Economic Research, Osaka University,
Department of System Neuroscience, School of Medicine, Sapporo Medical University

Makoto Kato
Kobe Advanced ICT Research Center

Kenju Akai
JSPS Research Fellow/ISER, Osaka University

Takuma Wakayama
Faculty of Economics, Ryukoku University

Tsutomu Murata
Kobe Advanced ICT Research Center

Tatsuyoshi Saijo
Institute of Social and Economic Research, Osaka University

May, 2009

Grant-in-Aid for Scientific Research on Priority Areas

The Cognitive Processing of Altruistic and Spiteful Behavior: An fMRI Study

Tsuyoshi Nihonsugi[1, 2]
Takafumi Yamakawa[1]
Jun Shinozaki[1, 3]
Makoto Kato[4]
Kenju Akai[2]
Takuma Wakayama[5]
Tsutomu Murata[4]
and
Tatsuyoshi Saijo[1]

May, 2009

[1] Institute of Social and Economic Research, Osaka University, 6-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan

[2] JSPS Research Fellow/ISER, Osaka University, 6-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan

[3] Department of System Neuroscience, School of Medicine, Sapporo Medical University, South 1 West 17, Chuo-ku, Sapporo, Hokkaido 060-8556, Japan

[4] Kobe Advanced ICT Research Center, 588-2, Iwaoka, Nishi-ku, Kobe, Hyogo 651-2492, Japan

[5] Faculty of Economics, Ryukoku University, 67 Tsukamoto-cho, Fukakusa, Fushimi-ku, Kyoto 612-8577, Japan

Corresponding author:

Tsuyoshi Nihonsugi
Osaka University
Institute of Social and Economic Research
6-1, Mihogaoka, Ibaraki,
Osaka 567-0047, Japan
+81-6-6879-8582 (voice)
+81-6-6878-2766 (fax)
E-mail: fge015nt@mail2.econ.osaka-u.ac.jp

ABSTRACT

The motivation underlying social behavior in humans, such as altruistic and spiteful behavior, is an important issue in the social sciences, and remains poorly understood. We used functional magnetic resonance imaging (fMRI) to investigate how humans evaluate the altruistic and spiteful behavior of others. Each subject inside the fMRI scanner was paired with a subject outside the scanner, and the two subjects played a reciprocal interactive game. Activation in posterior region of the rostral medial frontal cortex correlated with the detection of conflict while a subject was the recipient of altruistism. On the other hand, no areas were more active during the perception of spite. These findings suggest that subjects experienced a sense of conflict regarding their partner's altruistic behavior, indicating that humans may be suspicious when they are recipients of altruistic behavior that has no apparent purpose. This suspicion may be the underlying motivation of positive reciprocity.

Keywords: fMRI, experiment, altruism, spite, reciprocity, cognition
JEL Classification: C90, D87, H41

1 Introduction

Social behavior in humans, such as altruistic and spiteful behavior, has been studied not only in economics, but also in ethology, evolutionary psychology, neuroscience, and so on. The underlying motivation for and evolution of social behavior is an important question in the social sciences, and remains poorly understood (e.g., see Andreoni, 1995; Andreoni et al., 2007; Cason et al., 2004; Henrich, 2006; Houser and Kurzban, 2002; Nowak, 2006). Many studies indicate that these social behaviors are motivated by reciprocity. For example, Charness and Rabin (2002) conducted a range of simple experimental games (e.g., the Dictator game and Ultimatum game) and concluded that subjects are motivated by (positive or negative) reciprocity.

The promotion of reciprocal behavior is necessary so that humans are able to make inferences about the behavioral intentions of others. Therefore, in order to understand social behavior, it is important to understand how humans evaluate the behavior of others. In economics, very few studies are available that analyze how humans evaluate social behavior, such as altruism and spite. It is impossible to ask subjects to evaluate both their conscious and subconscious impressions of a partner's behavior, and only their conscious impressions can be understood using a questionnaire-based approach. To overcome this problem, we used functional magnetic resonance imaging (fMRI), which measures neural activation, to study both conscious and subconscious impressions while subjects evaluated the social behavior of others.

Recently, the ability to make inferences about the behavioral intentions and mental states of others has become widely known as "theory of mind" (Premack and Woodruff, 1978). Although many neuroscience studies have clarified which brain areas underlie theory of mind, there are few studies which shed light on the economic behavioral intentions of others in social interactions with real outcomes (Brunet et al., 2000; Fletcher et al., 1995; Gallagher et al., 2000; Saxe and Kanwisher, 2003; Vogeley et al., 2001).[1] For example, Sanfey et al. (2003) found that receiving unfair offers in ultimatum games differentially activated the anterior insula (a region involved in emotion), the dorsolateral prefrontal cortex (goal maintenance and executive control), and the anterior cingulate (detection of cognitive conflict). These findings suggest that subjects were disgusted by not only the amount of money offered, but the intentions of their partner, and this feeling of disgust led to the rejection of the offer. Rilling et al. (2004) found that several brain areas that had not previously been reported for theory of mind tasks were more significantly activated by human partners compared to computer partners in ultimatum games and sequential Prisoner's Dilemma Games.[2] However, this study did not focus on which particular aspects of the cognition of economic behavioral intentions correlate with different brain regions. Here, we used fMRI to investigate the neural substrates of the

---

[1] Gallagher and Frith (2003) describe the network of brain regions involved in theory of mind as comprising three main areas: the anterior paracingulate cortex, the posterior superior temporal sulcus, and the temporal pole.

[2] The differentially active regions included the posterior cingulate and precuneus, the mid superior temporal sulcus, the hypothalamus, the midbrain, the thalamus, and the hippocampus.

evaluation of altruistic and spiteful behavior during reciprocal interactions using simple payoff tables.

This study had two main goals: first, to determine the brain areas that correlate with evaluating the altruistic and spiteful behaviors of others; and second, to use the fMRI data to clarify the subconscious motivation underlying reciprocal behavior.

We observed activation in several areas that correlated with the evaluation of altruistic intention: posterior region of the rostral medial frontal cortex (prMFC), bilateral putamen, left anterior insula, right thalamus, and the parietal lobe. We suggest that the altruistic behavior of their partners were in conflict with subjects' expectations; for example, subjects might be suspicious of altruistic behavior when there is no apparent reason for the altruism. We concluded that this suspicion of altruistic behavior may be the subconscious motivation of positive reciprocity, as subjects opt to return the favor to err on the side of caution in social interactions.

The current study proceeds as follows: Section 2 presents the materials and methods of the experiment. Section 3 presents fMRI image acquisition and analysis. Section 4 presents the results. In section 5, the results are discussed. Section 6 draws conclusions about the results.

2 Materials and Methods

2.1 Subjects

Eighty-eight healthy right-handed male subjects were recruited from Osaka University by advertising leaflets. Written informed consent was obtained from all subjects. In the current study, four subjects were grouped together, and one of the four subjects in each group played the game in the fMRI scanner. Therefore, 22 of the 88 subjects underwent fMRI scanning. However, 6 of these 22 subjects were excluded from the analyses, because 2 subjects reported doubts about the sincerity of their partner's role, and 4 subjects had excessive head movement during scanning. Consequently, we analyzed data from 16 subjects (mean age of 20.9 years, SD = 1.9 years) for the group analysis.[3]

2.2 Experimental Design

Subjects were arranged into groups of 4 subjects who were unacquainted with each other. Subjects were provided with both written and pre-recorded oral instructions that explained the rules of the game.

After the instructions, subjects were given 10 minutes to ask questions, after which they were tested to confirm that they understood the rules. Then, 1 of the 4 subjects was selected by lottery to play the game in the fMRI scanner (this subject is called the "insider"). The other three subjects

---

[3] Sixteen subjects is an adequate sample size for an fMRI study.

played the game in a room adjacent to the fMRI scanner (the “outsiders”). Each outsider was paired with the insider, and the game was played in an interactive manner using a payoff table. Paired subjects viewed the same payoff tables on their screens. The subjects who were not playing the game were not allowed to watch the screen, and could not confer with each other.

Each of the outsiders was assigned a role in the game. The first subject, “outsider X”, was asked to behave in a way that maximized their own payoff; “outsider Y” was required to behave in a spiteful manner; and “outsider Z” was asked to behave altruistically. The insider did not know that the other subjects were assigned certain roles.[4]

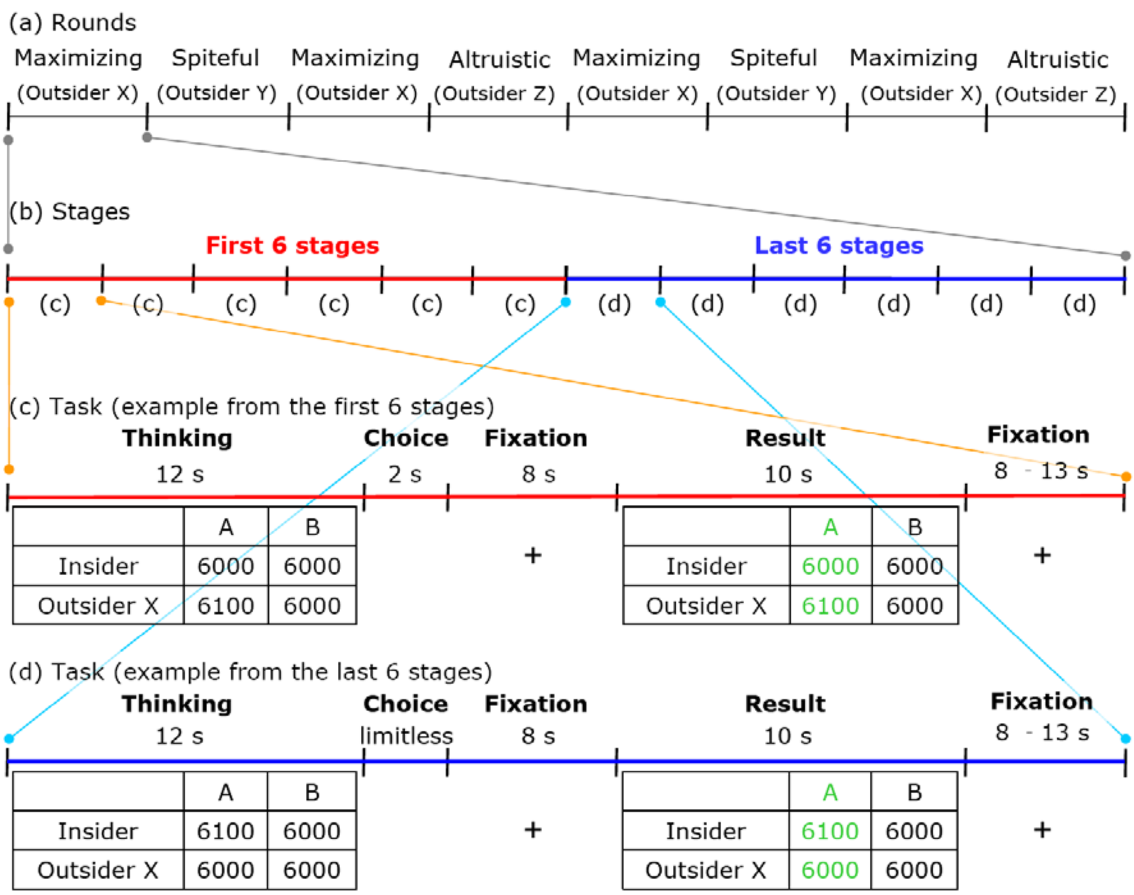

Fig. 1 Schematic of the experiment.

Figure 1a shows an example timeline of the experiment. The game consisted of eight rounds: four Maximizing Rounds in which the insider was paired with outsider X, two Spiteful Rounds in which the insider was paired with outsider Y, and two Altruistic Rounds in which the insider was paired

[4] A deceptive scenario was used in the current study because of the heavy logistical demands of an fMRI study. Therefore, the two subjects who reported doubts about the sincerity of their partner’s role in the game were excluded from the analysis.

with outsider Z. To control for the effects of task order, four different order patterns were used.[5]

Each round consisted of 12 stages (Fig. 1b). The first 6 stages were performed continuously, during which one of the outsiders made a decision, while the insider only watched. The last 6 stages were performed continuously, and during these stages the insider made a decision, while the outsiders observed his actions.

Figure 1c shows the timeline for one of the first 6 stages. Each stage began with the screens showing a payoff table for 12 seconds (Thinking). When the letters A and B turned blue, an outsider pressed one of two buttons, choosing between A or B (Choice). For example, when outsider X chose A, the insider received 6,000 points and outsider X received 6,100 points. On the other hand, when outsider X chose B, the insider received 6,000 points and outsider X received 6,000 points.[6] Then, a fixation cross was presented for 8 seconds (Fixation).[7] After that, the result chosen by the outsider appeared in green for 10 seconds (Result). A fixation cross was then presented for a variable time period of 8 to 13 seconds (Fixation).

Fig. 1d shows the timeline for one of the last 6 stages. Each stage began with a payoff table appearing for 12 seconds (Thinking). When the letters A and B became blue, the insider pressed one of two buttons to choose between A or B (Choice). After this, a fixation cross was presented for 8 seconds (Fixation), followed by the appearance of the insider's choice in green for 10 seconds (Result). Then, a fixation cross was presented for a variable time period of 8 to 13 seconds (Fixation).

Payment to the subjects was proportional to the number of points earned during the experiment. The mean payoff per subject was 15,000 Yen (about $150), and the experiment required six hours to complete, including three hours of transportation time.

2.3 Payoff tables

We used three types of payoff tables. Figs. 2a, 2b and 2c provide examples of the payoff tables used during the Maximizing Round, the Spiteful Round and the Altruistic Round, respectively. The payoffs in each table changed in 100-point steps (see Appendix), but the general appearance of the payoff tables was not altered.

	A	B
Insider	6000	6000
Outsider X	6100	6000

Fig. 2a An example of a payoff table during the Maximizing Round

[5] The four order patterns were: (1) M → S → M → A → M → S → M → A, (2) M → A → M → S → M → A → M → S, (3) M → S → M → S → M → A → M → A, (4) M → A → M → A → M → S → M → S. M: maximizing round, S: spiteful round, A: altruistic round.

[6] 1,000 experimental points = 40 Yen (about $0.4).

[7] The "fixation cross" appears in the center of a blank screen in order to hold the visual attention of subjects.

	A	B
Insider	2300	6100
Outsider Y	6000	6100

Fig. 2b An example of a payoff table during the Spiteful Round

	A	B
Insider	9300	6100
Outsider Z	6000	6100

Fig. 2c An example of a payoff table during the Altruistic Round

The payoff tables above were used when one of the outsiders chose between A or B. The defining feature of the Maximizing Round (Fig. 2a) was that no matter which choice outsider X made, the insider's payoff did not change. Therefore, in the Maximizing Round there was no interaction between the players, and each subject considered only his own payoff. The defining feature of the Spiteful Round (Fig. 2b) was that outsider Y could drastically reduce the payoff to the insider, while reducing their own payoff only slightly (spiteful behavior). Therefore, the Spiteful Round involved negative reciprocity. The defining feature of the Altruistic Round (Fig. 2c) was that outsider Z could drastically increase the payoff to the insider, while reducing their own payoff only slightly (altruistic behavior). Therefore, the Altruistic Round entailed positive reciprocity.

When the insider chose A or B, the payoffs were reversed compared with when one of the outsiders chose A or B (see Appendix).

3 fMRI Image Acquisition and Analysis

3.1 fMRI Image Acquisition

Functional images were acquired on a Siemens 3T Trio scanner using an echo planar imaging (EPI) sequence with the following parameters: repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, matrix = 64 × 64, field of view (FOV) = 192 mm, slice thickness = 3 mm, gap = 0 mm, interleaved slice acquisition of 34 axial slices, yielding 3-mm cubic voxels. T1-weighted magnetization prepared rapid gradient echo (MPRAGE) anatomical scans were also acquired (TR = 2000 ms, TE = 4.38 ms, matrix = 256 × 256, FOV=230mm, slice thickness = 1 mm).

3.2 fMRI Image Analysis

Preprocessing of functional imaging data was performed using Statistical Parametric Mapping 5 (SPM5, www.fil.ion.ucl.ac.uk/spm/). Preprocessing included slice-timing correction (centered at

TR/2), motion correction, coregistration to the individual subjects' anatomical images, spatial normalization to the Montreal Neurological Institute (MNI) template, and spatial smoothing using an 8 mm Gaussian kernel. The data were then detrended using a high-pass filter of periods greater than 128 seconds.

Statistical analysis involved a simple linear regression where dummy variables were "on" when a payoff table was on the screen and "off" otherwise. This "boxcar" regression was convolved with the hemodynamic response function. The regression coefficients of activity in the blood oxygen-dependent level (BOLD) signal in each voxel indicated which voxels were significantly active.

We defined nine regressors: Outsider's Thinking (duration = 12 s); Outsider's Thinking (duration = 1 s); Outsider's Result (duration = 2 s)[8]; Outsider's Result (duration = 10 s); Outsider's Thinking (duration = 3 s)[8]; Insider's Thinking + Insider's Choice (duration = 12 s + reaction time); Insider's Choice (duration = reaction time); Insider's Result (duration = 1 s); Insider's Result (duration = 10 s).

The resulting general linear model was corrected for temporal autocorrelation using a first-order autoregressive model. These individual contrast images were then submitted to a second-level random-effects analysis. Statistical maps were thresholded for significance ($p < 0.001$, uncorrected) and cluster size ($\geqq$ 110 voxels). The MNI coordinates were converted to Talairach coordinates (Talairach and Tournoux, 1988).

4 Imaging Results

As mentioned above, it is important to analyze not only the activity related to decision-making but also the activity related to the evaluation of the decision-making of other participants. Therefore, we focused our analysis on the neural substrates of the cognitive and emotional processes (i.e., the Outsider's Result [duration = 2 s]) involved in being the recipient of altruistic and spiteful behavior compared with being the recipient of maximizing behavior.

4.1 Receiving altruistic vs. Maximizing behavior

We identified the brain areas that were correlated with the altruistic intentions of others by contrasting the activity related to receiving altruistic behavior and the activity related to receiving maximizing behavior. This contrast yielded significant activations in posterior region of the rostral

[8] In additional behavioral experiments, we found that subjects were able to recognize the outsider's choice in 2 seconds (mean reaction time: 2.08 seconds, SD = 1.18 seconds) and that the subject was able make their own choice in 3 seconds (mean reaction time: 3.07 seconds, SD = 1.52 seconds). Therefore, in the current study we chose 2 second and 3 second time-windows to analyze the data.

medial frontal cortex (prMFC, Brodmann area 6),[9] bilateral putamen, left anterior insula, right thalamus, and the parietal lobe (Brodmann area 5) (Fig. 3, Table 1).

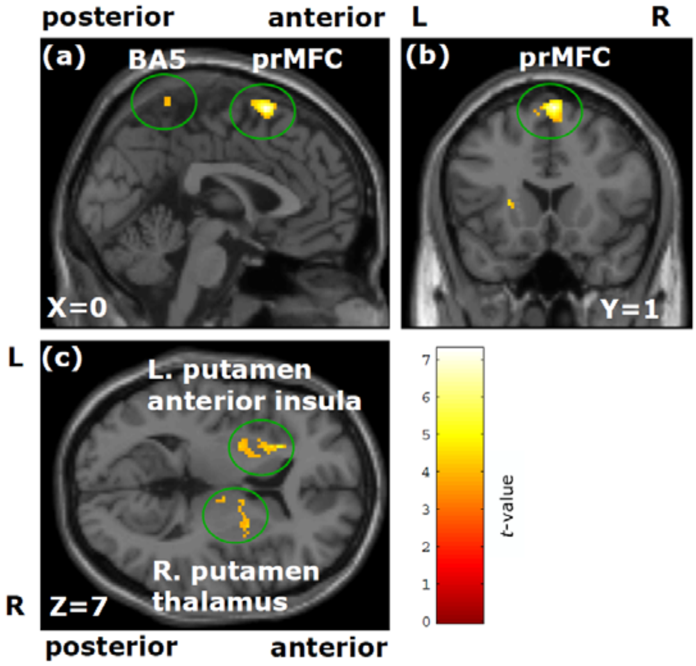

Fig. 3 Colored maps of the t statistic for the contrast between the trials in which the subject was the recipient of a partner's altruistic behavior and those in which the partner exhibited maximizing behavior. Maps are thresholded at $p < 0.001$, uncorrected. The three panels each show the brain activity using a color scale to indicate the level of statistical significance. A 3-dimensional coordinate system is used which locates the anterior commissure at $x = y = z = 0$. BA: Brodmann area; prMFC: posterior region of the rostral medial frontal cortex; R: right, L: left.

Table 1. Areas activated when receiving a partner's altruistic behaviors compared to receiving a partner's maximizing behaviors.

Region	Talairach coordinates (x y z)	Voxel size (voxels)	t value
prMFC (pre-SMA)	-2 17 60	247	7.27
Left putamen/anterior insula	-24 22 4	133	5.17
Right putamen /thalamus	24 1 9	228	5.1
Parietal lobe (BA 5)	16 -44 61	159	5.01

Note: Talairach coordinates (x, y, z) indicate the location of the maximum changes in BOLD signal. The threshold is set at an uncorrected $p < 0.001$. The voxel size shows the number of supra-threshold voxels, and the t values represent the score for the peak activation voxel.

[9] Amodio and Frith (2006) refer to this area as the prMFC, which includes the presupplementary motor area (pre-SMA, BA6) and the dorsal anterior cingulate cortex.

Next, we focused on the prMFC activation. The mean activities in the prMFC (x, y, z coordinates -2, 17, 60) of all subjects are shown in Fig. 4. We conducted an ANOVA with one between-subject factor (subject) and one within-subject factor (round type) to analyze the prMFC activity. There was a significant main effect of round type (F (2,15) = 10.60, $p < 0.001$). A *post hoc* Tukey test revealed a significant difference between the strength of neural activation when subjects were recipients of altruistic behavior and the strength of neural activation when the subjects were recipients of maximizing behavior ($p < 0.001$).

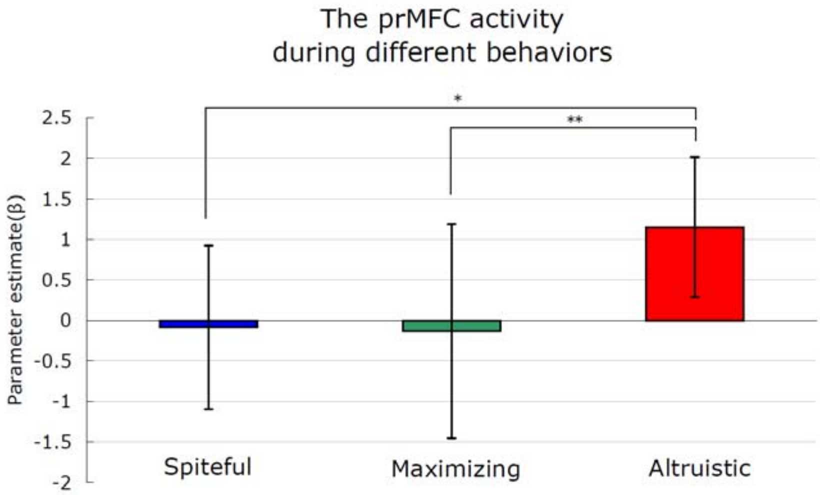

Fig. 4 Illustration of the strength of the average activation (parameter estimates) in the prMFC (-2, 17, 60) when subjects were recipients of spiteful, maximizing and altruistic behavior compared with fixation. Error bars indicate the standard error of the mean (SEM). ** $p < 0.001$, * $p < 0.002$ (Tukey test).

In the reverse contrast (i.e., Receiving Maximizing Behavior vs. Receiving Altruistic Behavior), the whole-brain analysis did not reveal any significant activation.

4.2 Receiving spiteful behavior vs. Maximizing behavior

We identified the brain areas that correlated with the spiteful intentions of others. This contrast did not yield significant activation in the whole-brain analysis. Similarly, the reverse contrast (i.e., Receiving Maximizing Behavior vs. Receiving Spiteful Behavior), did not reveal any significant activation in the whole-brain analysis.

5 Discussion

Our objective in the current study was to investigate the neural substrates of evaluating the altruistic and spiteful intentions of others. As previously mentioned, earlier studies have not shed light on the neural mechanisms of evaluating other’s economic behaviors. Activation in several areas correlated with the evaluation of altruistic intentions, including posterior region of the rostral medial frontal cortex (prMFC), bilateral putamen, left anterior insula, right thalamus, and the parietal lobe.

Several previous studies suggest that the prMFC is involved in the detection of both negative feedback (e.g., conflict and errors) and positive feedback (e.g., correct responses). For example, Ruff et al. (2001) observed activity in the prMFC during different forms and degrees of conflict between the word and the color dimensions of Stroop stimuli.[10] They suggested that the prMFC is differentially sensitive to various type of conflict.[11] Similar results were found using conflict tasks other than the Stroop task (e.g., Brázdil et al., 2005; Milham et al., 2003; Pochon et al., 2008; Ullsperger and von Cramon, 2003; Ullsperger and von Cramon, 2004).[12] When we asked subjects how they felt about their partner's behavior, certain subjects indicated that they had not predicted that their partners would behave altruistically, but instead anticipated spiteful behavior, suggesting that there was a conflict between their expectations and their experiences during the altruism condition.

Also, several previous studies suggest that the prMFC is involved in the assessment of the outcomes of both one's own and other's choices in order to better guide subsequent decisions. For example, Walton et al. (2004) observed activity in the prMFC when subjects monitored the outcome of actions that were self-selected. They suggested that the prMFC is involved in assessing the value of chosen responses and guiding subsequent choices accordingly.[13] Coricelli et al. (2005) conducted a study in which subjects selected between two possible gambles, and regret was induced by providing information about the higher reward outcome of the gamble that was not chosen. They observed that activity in the prMFC increased with increasing degrees of regret, and they also concluded that the prMFC is involved in improved guidance of subsequent choices.[14] Another study found that the prMFC is involved in not only the assessment of self-generated error but also the assessment of errors generated by others (van Schie et al. 2004).[15] Taken together, these findings imply that the prMFC activity is related to the assessment of other's altruistic intentions in complex social scenarios.

Putamen activation is often observed in economic decision-making tasks. Satpute and Lieberman (2006) propose that the putamen is involved in the evaluation and prediction of reward.

We also investigated the neural substrates involved in evaluating the spiteful behavior of others. We did not observe any area that showed greater activation in response to a partner's spiteful

---

[10] For example, in the Stroop color-naming task, subjects viewed the word "blue" written in green letters, and were asked "What color is this?". Ruff et al. (2001) used more complicated Stroop stimuli.

[11] Ruff et al. (2001) observed increased activity at x, y, z coordinates (-8, 18, 47) and (-4, 7, 55).

[12] Milham et al. (2003) reported increased activity at x, y, z coordinates (0, 20, 46) and (0, 10, 52); Ullsperger and Cramon (2003) observed increased activity at (0, 13, 53), and Ullsperger and Cramon (2004) at (4, 18, 53); Brázdil et al. (2005) at (3, 25, 40) and (27, 11, 52); and Pochon et al. (2008) found increased activity at (-3, 20, 43).

[13] Walton et al. (2004) observed increased activity at x, y, z coordinates (0, 12, 36).

[14] Coricelli et al. (2005) found increased activity at x, y, z coordinates (10, 25, 30).

[15] Schie et al. (2004) observed increased activity at x, y, z coordinates (4, 5, 22).

behavior compared to a partner's maximizing behavior. Although the small number of rounds in which spiteful behavior occurred (n = 2) could limit our statistical power to detect significant activity, one possible reason for this result may be that subjects accurately predicted that their partner would behave in a spiteful or maximizing way, and considered this behavior as natural.

6. Conclusions

The main result from the imaging data was increased activity in the prMFC, suggesting that subjects were suspicious of the altruistic behavior of others, as this behavior conflicted with what the subjects anticipated of their partners in the game. Furthermore, we suggest that the motivation for positive reciprocity is based on the suspicion aroused by unexpected altruistic behavior, and that humans opt to for positive reciprocity as a self-protective measure.

The finding that there was no differential activity when subjects were evaluating a partner's spiteful behavior indicates that, in this condition, there was no conflict between the subject's expectation and their partner's behavior.

In ethology, it is generally known that behavioral optimization is made up of two complementary processes: evaluating the outcomes of an executed behavior (performance monitoring) and adjusting subsequent behavior based on these outcomes (performance adjustment). Humans usually recognize the behavioral intentions of others and adjust their subsequent behavior accordingly. This learning process (performance monitoring and adjustment) is important to adjust to a changing environment, and occurs during reciprocal interactions in humans. Our findings indicate that one underlying motive of positive reciprocity is the suspicion of the intentions of others.

Acknowledgments

We appreciate the advice and expertise of Seigi Itoi, Akitoshi Seiyama, Sachiko Takahama and Yasuto Tanaka. We also thank the seminar participants at Osaka University, Tamagawa University, the 2007 Asia Pacific Regional Meeting of Economic Science Association, and the 2007 Japanese Economic Association Autumn Meeting at Nihon University for comments and discussions. We are also grateful to the Kobe Advanced ICT Research Center for allowing us to use the fMRI scanner. This research was supported by the 21st century COE program (Osaka University).

References

Amodio, D. M., Frith, C. D., 2006. Meeting of minds: the medial prefrontal cortex and social cognition. Nature Reviews Neuroscience. 7, 268–277.

Andreoni, J., 1995. Cooperation in Public Goods Experiments: Kindness or Confusion? American Economic Review. 85, 891–904.

Andreoni, J., Harbaugh, W. T., Vesterlund, L., 2007. Altruism in Experiments. In: Durlauf, S. N., Blume, L. E. (Eds.), The New Palgrave Dictionary of Economics, 2nd edition. United Kingdom: Palgrave Macmillan.

Brázdil, M., Dobsik, M., Mikl, M., Hlustik, P., Daniel, P., Pazourkova, M., Krupa, P., Rektor, I., 2005. Combined event-related fMRI and intracerebral ERP study of an auditory oddball task. NeuroImage. 26, 285–293.

Brunet, E., Sarfati, Y., Hardy-Bayle, M. -C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. NeuroImage. 11, 157–166.

Cason, T., Saijo, T., Yamato, T., Yokotani, K, 2004. Non-excludable public good experiments. Games and Economic Behavior. 49, 81–102.

Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics. 117, 817–869.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., Dolan, R. J., 2005. Regret and its avoidance: a neuroimaging study of choice behavior. Nature Neuroscience. 8, 1255–1262.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., Frith, C. D., 1995. Other minds in the brain: a functional imaging study of ‘‘theory of mind’’ in story comprehension. Cognition. 57, 109–128.

Gallagher, H., Frith, C., 2003. Functional imaging of ‘‘theory of mind’’. Trends in Cognitive Science. 7, 77–83.

Gallagher, H., Happe, F., Brunswick, N., Fletcher, P., Frith, U., Frith, C., 2000. Reading the mind in cartoons and stories: an fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. Neuropsychologia. 38, 11 –21.

Henrich, J., 2006. Cooperation, punishment, and the evolution of human institutions. Science. 312, 60–61.

Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments. The American Economic Review. 92, 1062–1069.

Milham, M. P., Banich, M. T., Claus, E. D., Cohen, N. J., 2003. Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. NeuroImage. 18, 483–493.

Nowak, M. A., 2006. Five rules for the evolution of cooperation. Science. 314, 1560–1563.

Pochon, J. B., Riis, J., Sanfey, A. G., Nystrom, L. E., Cohen, J. D., 2008. Functional imaging of decision conflict. Journal of Neuroscience. 28, 3468–3473.

Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences. 1, 515–526.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., Cohen, J. D., 2004. The neural correlates of Theory of Mind within interpersonal interactions. NeuroImage. 22, 1694–1703.

Ruff, C. C., Woodward, T. S., Laurens, K. R., Liddle, P. F., 2001. The role of the anterior cingulate cortex in conflict processing: evidence from reverse Stroop interference. NeuroImage. 14, 1150–1158.

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, 1755–1758.

Satpute, A. B., Lieberman, M. D., 2006. Integrating automatic and controlled processes neurocognitive models of social cognition. Brain Research. 1079, 86–97.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". NeuroImage. 19, 1835–1842.

Talairach, J., & Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. New York: Thieme Medical Publishers.

Ullsperger, M., von Cramon, D. Y., 2003. Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. Journal of Neuroscience. 23, 4308–4314.

Ullsperger, M., von Cramon, D. Y., 2004. Neuroimaging of performance monitoring: error detection and beyond. Cortex. 40, 593–604.

van Schie, H. T., Mars, R. B., Coles, M. G. H., Bekkering, H., 2004. Modulation of activity in medial frontal and motor cortices during error observation. Nature Neuroscience. 7, 549–554.

Vogeley, K., Bussfeld, P., Newen, A., Hermann, S., Happe, F., Falkai, P., Maier, W., Shah, N., Fink, G., Zilles, K., 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. NeuroImage. 14, 170–181.

Walton, M. E., Devlin, J. T., Rushworth, M. F., 2004. Interactions between decision making and performance monitoring within prefrontal cortex. Nature Neuroscience. 7, 1259–1265.

Supplementary materials

These supplementary materials report the details of several contrasts briefly mentioned in the main text.

1 The payoff tables used when an outsider chose between A or B

Maximizing Round

	A	B
Insider	6000	6000
Outsider (X)	5800	5300

	A	B
Insider	6000	6000
Outsider (X)	5800	5400

	A	B
Insider	6000	6000
Outsider (X)	5800	5500

	A	B
Insider	6000	6000
Outsider (X)	5800	5700

	A	B
Insider	6000	6000
Outsider (X)	5400	5900

	A	B
Insider	6000	6000
Outsider (X)	5900	5500

	A	B
Insider	6000	6000
Outsider (X)	5900	5700

	A	B
Insider	6000	6000
Outsider (X)	5900	5800

	A	B
Insider	6000	6000
Outsider (X)	5500	6000

	A	B
Insider	6000	6000
Outsider (X)	6000	5700

	A	B
Insider	6000	6000
Outsider (X)	5600	6100

	A	B
Insider	6000	6000
Outsider (X)	6100	5700

	A	B
Insider	6000	6000
Outsider (X)	5800	6100

	A	B
Insider	6000	6000
Outsider (X)	6100	6000

	A	B
Insider	6000	6000
Outsider (X)	5800	6200

Spiteful Round

	A	B
Insider	5800	2300
Outsider (Y)	5800	5700

	A	B
Insider	5900	2300
Outsider (Y)	5900	5800

	A	B
Insider	6000	2600
Outsider (Y)	6000	5900

	A	B
Insider	2300	6100
Outsider (Y)	6000	6100

	A	B
Insider	2400	6100
Outsider (Y)	6000	6100

	A	B
Insider	2500	6200
Outsider (Y)	6100	6200

	A	B
Insider	6200	2600
Outsider (Y)	6200	6100

	A	B
Insider	6200	2700
Outsider (Y)	6200	6100

Altruistic Round

	A	B
Insider	5800	9700
Outsider (Z)	5800	5700

	A	B
Insider	5900	9300
Outsider (Z)	5900	5800

	A	B
Insider	6000	9600
Outsider (Z)	6000	5900

	A	B
Insider	9700	6000
Outsider (Z)	5900	6000

	A	B
Insider	9300	6100
Outsider (Z)	6000	6100

	A	B
Insider	9600	6100
Outsider (Z)	6000	6100

	A	B
Insider	6100	9700
Outsider (Z)	6100	6000

	A	B
Insider	6200	9300
Outsider (Z)	6200	6100

	A	B
Insider	6200	9400
Outsider (Z)	6200	6100

	A	B
Insider	6200	9500
Outsider (Z)	6200	6100

2 The payoff tables used when the insider chose between A or B

Maximizing Round

	A	B
Insider	5300	5800
Outsider (X)	6000	6000

	A	B
Insider	5500	5800
Outsider (X)	6000	6000

	A	B
Insider	5800	5400
Outsider (X)	6000	6000

	A	B
Insider	5900	5700
Outsider (X)	6000	6000

	A	B
Insider	5900	5800
Outsider (X)	6000	6000

	A	B
Insider	6000	5600
Outsider (X)	6000	6000

	A	B
Insider	5700	6000
Outsider (X)	6000	6000

	A	B
Insider	6000	5800
Outsider (X)	6000	6000

	A	B
Insider	6000	5900
Outsider (X)	6000	6000

	A	B
Insider	6100	5600
Outsider (X)	6000	6000

	A	B
Insider	6100	5700
Outsider (X)	6000	6000

	A	B
Insider	5800	6100
Outsider (X)	6000	6000

	A	B
Insider	6100	6000
Outsider (X)	6000	6000

	A	B
Insider	6200	5700
Outsider (X)	6000	6000

	A	B
Insider	6200	5800
Outsider (X)	6000	6000

	A	B
Insider	6200	5900
Outsider (X)	6000	6000

	A	B
Insider	6100	6200
Outsider (X)	6000	6000

Spiteful Round

	A	B
Insider	5800	2300
Outsider (Y)	5800	5700

	A	B
Insider	5900	2300
Outsider (Y)	5900	5800

	A	B
Insider	6000	2600
Outsider (Y)	6000	5900

	A	B
Insider	2300	6100
Outsider (Y)	6000	6100

	A	B
Insider	2400	6100
Outsider (Y)	6000	6100

	A	B
Insider	2500	6200
Outsider (Y)	6100	6200

	A	B
Insider	6200	2600
Outsider (Y)	6200	6100

	A	B
Insider	6200	2700
Outsider (Y)	6200	6100

Altruistic Round

	A	B
Insider	5800	5700
Outsider (Z)	5800	9300

	A	B
Insider	5800	5700
Outsider (Z)	5800	9500

	A	B
Insider	5700	5800
Outsider (Z)	9600	5800

	A	B
Insider	5900	5800
Outsider (Z)	5900	9300

	A	B
Insider	5800	5900
Outsider (Z)	9400	5900

	A	B
Insider	6000	5900
Outsider (Z)	6000	9500

	A	B
Insider	5900	6000
Outsider (Z)	9700	6000

	A	B
Insider	6000	6100
Outsider (Z)	9700	6100

	A	B
Insider	6200	6100
Outsider (Z)	6200	9500

	A	B
Insider	6100	6200
Outsider (Z)	9700	6200

3 Additional Results

3.1 Receiving altruistic behavior vs. spiteful behavior

No areas were significantly activated in the whole-brain analysis of either this contrast or the reverse contrast (i.e., Received Altruistic Behavior vs. Received Spiteful Behavior).

3.2 (Received Altruistic Behavior + Spiteful Behavior) vs. Received Maximizing Behavior

We identified brain areas that correlated with social interaction (so-called theory of mind) by comparing the time periods when subjects were the recipients of a partner's altruistic and spiteful behavior vs. when subjects observed a partner's maximizing behavior. When we tested this hypothesis, we found that no areas were activated in the whole-brain analysis or in the reverse contrast (i.e., Received Maximizing Behavior vs. [Received Altruistic Behavior + Spiteful Behavior]).