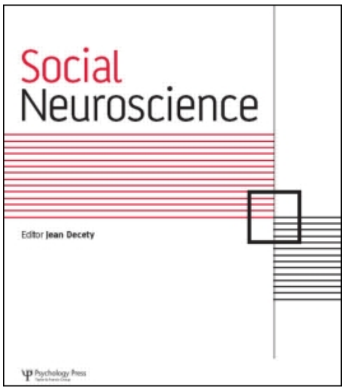

Social Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t741771143

The roles of the medial prefrontal cortex and striatum in reputation processing

Keise Izuma [ab]; Daisuke N. Saito [acd]; Norihiro Sadato [aefg]
[a] National Institute for Physiological Sciences, Aichi [b] Graduate University for Advanced Studies, Kanagawa, Japan [c] Japan Science and Technology Agency (JST)/Research Institute of Science, [d] Technology for Society (RISTEX), Saitama, Japan [e] Graduate University for Advanced Studies, Kanagawa [f] JST/RISTEX, Saitama [g] University of Fukui, Fukui, Japan

First Published on: 24 August 2009

To cite this Article Izuma, Keise, Saito, Daisuke N. and Sadato, Norihiro(2009)'The roles of the medial prefrontal cortex and striatum in reputation processing',Social Neuroscience,99999:1,

To link to this Article: DOI: 10.1080/17470910903202559

URL: http://dx.doi.org/10.1080/17470910903202559

The roles of the medial prefrontal cortex and striatum in reputation processing

Keise Izuma

National Institute for Physiological Sciences, Aichi, and Graduate University for Advanced Studies, Kanagawa, Japan

Daisuke N. Saito

National Institute for Physiological Sciences, Aichi, and Japan Science and Technology Agency (JST)/Research Institute of Science and Technology for Society (RISTEX), Saitama, Japan

Norihiro Sadato

National Institute for Physiological Sciences, Aichi, and Graduate University for Advanced Studies, Kanagawa, and JST/RISTEX, Saitama, and University of Fukui, Fukui, Japan

How we are viewed by other individuals—our reputation—has a considerable influence on our everyday behaviors and is considered an important concept in explaining altruism, a uniquely human trait. Previously it has been proposed that processing one's own reputation requires a reputation representation in the medial prefrontal cortex (mPFC) and a value representation in the striatum. Here, we directly tested this idea using functional magnetic resonance imaging (fMRI). Subjects disclosed their behavioral tendencies with reference to social norms in the presence or absence of other people, a manipulation that is known to greatly affect an individual's concern for their reputation. The mPFC showed strong activation during self-referential processing, and this activity was enhanced by the mere presence of observers. Moreover, the striatum was also strongly activated when subjects responded in front of observers. Thus, the present study demonstrated that the mPFC and striatum were automatically recruited when the task placed a high demand on processing how one is viewed by others. Taken together, our findings suggest that the mPFC and the striatum play a key role in regulating human social behaviors, and these results provide valuable insight into the neural basis of human altruism.

Keywords: mPFC; Striatum; Reputation; Social reward.

INTRODUCTION

An individual's reputation is an evaluation made by other people with regard to socially desirable or undesirable behaviors, and can be thought of as a "meta-belief" (i.e., a belief about how others view us). Because it is an evaluation, a reputation always contains some reward value (either positive or negative), and in real-life social interactions the concern for a good reputation often affects how people behave in social settings. People care about how they are viewed by others,

Correspondence should be addressed to: N. Sadato, Section of Cerebral Research, Division of Cerebral Integration, National Institute for Physiological Sciences (NIPS), Okazaki, Aichi, 444-8585, Japan. E-mail: sadato@nips.ac.jp

We wish to thank M. J. Hayashi, Y. Mano, and A. Sasaki for their assistance in conducting the experiment, and M. Sugiura for his comments on the manuscript. This study was supported, in part, by a Grant-in-Aid for Scientific Research (S#17100003 and S#21220005 to NS) from the Japan Society for the Promotion of Science.

www.psypress.com/socialneuroscience DOI:10.1080/17470910903202559

and behave differently depending on whether or not others are present.

Theoretical research on the evolution of human cooperation also highlights the importance of one's reputation or how one is viewed by others. Social scientists and evolutionary biologists have long studied the question of why humans cooperate with genetically unrelated individuals, which seems to be a distinguishing feature of our species (Fehr & Fischbacher, 2003). A person's reputation is one of the key mechanisms explaining human altruism through indirect reciprocity (Nowak, 2006; Nowak & Sigmund, 2005). Experimental evidence suggests that humans are sensitive to the possibility of establishing a reputation (Fehr & Fischbacher, 2003), and that an individual's motivation to acquire a good reputation or "image score" (Milinski, Semmann, Bakker, & Krambeck, 2001; Nowak & Sigmund, 1998; Wedekind & Milinski, 2000) might drive cooperation through indirect reciprocity. Furthermore, the reputation mechanism helps to establish cooperation even in public good situations, through alternating rounds of public good and indirect reciprocity games (Milinski, Semmann, & Krambeck, 2002).

Consistent with theoretical expectations, social psychological studies also suggest that humans possess a strong drive to gain a good reputation or impression in the eyes of others, which works as an incentive for prosocial behaviors (Benabou & Tirole, 2006), and that even subtle cues indicative of being observed are sufficient to increase an individual's concern for their reputation. For example, subtle manipulations of anonymity (e.g., asking someone to provide their name when answering a questionnaire) affect an individual's tendency to respond in a socially desirable manner (i.e., the social desirability tendency) (Crowne & Marlowe, 1960; Lautenschlager & Flaherty, 1990; Nederhof, 1985; Paulhus, 1984). Similarly, subtle observation cues (i.e., pictures of eyes or eye-like stimuli) have been shown to be sufficient to enhance prosocial behaviors in both laboratory (Haley & Fessler, 2005; Kurzban, DeScioli, & O'Brien, 2007) and real-life situations (Bateson, Nettle, & Roberts, 2006).

Although the notion of reputation plays an important role in various fields of the social sciences, and there are many examples of behaviors affected by the concern for a positive social image in our everyday lives, how reputation is processed in human brains has not been fully investigated. We previously proposed that the neural basis of social reward processing (one's good reputation) consists of a reputation representation in the medial prefrontal cortex (mPFC) and a value representation in the striatum (Izuma, Saito, & Sadato, 2008). While a previous study demonstrated that the value associated with the social reward of a good reputation activated the striatum in a similar manner to monetary reward (Izuma et al., 2008), there was only preliminary evidence that the mPFC was involved in understanding one's reputation, due to the lack of appropriate control conditions. The mPFC's function of forming complex and abstract representations is thought to be essential (Amodio & Frith, 2006) to the ability to reflect on the value associated with a good reputation. Although previous studies (D'Argembeau et al., 2007; Ochsner et al., 2005) reported that mPFC is involved in viewing oneself from others' perspective, these studies provided only indirect evidence for mPFC's role in one's reputation processing because they manipulated the instruction given to the subjects and explicitly asked them to see themselves from the perspective of others.

In the present study, we sought to further test the roles played by the mPFC and the striatum in reputation processing. We did this by systematically and directly manipulating the situational factor of the presence/absence of other people, which has been shown to influence an individual's concern for his/her reputation (Bateson et al., 2006; Haley & Fessler, 2005; Kurzban et al., 2007), while subjects engaged in various judgment tasks (see below).

According to Amodio and Frith (2006), the representation of one's own reputation requires the formation of a second-level representation of the attributes that others assign to us (i.e., thinking about what others think of us), and the mPFC (also called the anterior rostral medial frontal cortex, arMFC) may play a crucial role in this process. The mPFC has been implicated in complex social cognitive processing in human neuroimaging studies using various cognitive tasks. The first line of research reporting robust activation in the mPFC includes studies using theory of mind or mentalizing tasks, in which subjects were required to represent another person's mental state (for review, see Gallagher & Frith, 2003). In these studies, mPFC activity was consistently observed during various experimental tasks, including a task that required subjects to infer another's knowledge about a certain tool (Goel, Grafman, Sadato, & Hallett, 1995), story, and/or cartoon comprehension tasks that required subjects to infer the mental state of another person (Brunet, Sarfati, Hardy-Bayle, & Decety, 2000; Fletcher et al., 1995; Gallagher et al., 2000), and animations of simple geometrical shapes that evoked mental state attribution (Castelli, Happé, Frith, & Frith, 2000). Mentalizing-related activity in the mPFC has been also observed in active online tasks. These studies have shown greater mPFC activity while subjects played interactive games, such as the Prisoner's Dilemma game, with human partners compared to computer opponents (Fukui et al., 2006; Gallagher, Jack, Roepstorff, & Frith, 2002; McCabe, Houser, Ryan, Smith, & Trouard, 2001; Rilling et al., 2002; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004).

The second line of evidence highlights mPFC activation during self-referential processing (for review, see Northoff & Bermpohl, 2004). In a typical self-reference task, subjects are required to judge whether a personality trait or a statement about attitudes accurately describes them. During the task, activity in the mPFC was significantly increased compared to the control task (when subjects determined whether the same word described another person) (D'Argembeau et al., 2007; Fossati et al., 2003; Johnson et al., 2002; Kelley et al., 2002; Ochsner et al., 2005; Schmitz, Kawahara-Baccus, & Johnson, 2004; Zysset, Huber, Ferstl, & von Cramon, 2002). Furthermore, the mPFC activation during self-reflection and theory of mind tasks largely overlapped at the individual subject level (Saxe, Moran, Scholz, & Gabrieli, 2006). Moreover, greater mPFC activation was reported not only when individuals were thinking about themselves, but also when subjects were thinking about the mental state of a similar other; on the contrary, thinking about unknown others activated a more dorsal part of the mPFC (Mitchell, Banaji, & Macrae, 2005). Based on the findings in these studies that employed a wide range of cognitive tasks, several researchers suggested that the mPFC (or arMFC) plays a crucial role in forming more complex and abstract representations, or metacognitive representations (i.e., thinking about thinking), which allow us to reflect on what other people think of us (Amodio & Frith, 2006; Ochsner et al., 2005). Also, another line of evidence assumes that the mPFC represents abstract dynamic summary representations as the underlying structures for the development of event, person and self schemata (Krueger, Barbey, & Grafman, 2009).

On the other hand, the striatum is a brain area known to be related to reward processing (for review, see Delgado, 2007; Schultz, Tremblay, & Hollerman, 2000). Both neuroimaging studies with humans (Delgado, Locke, Stenger, & Fiez, 2003; Knutson, Adams, Fong, & Hommer, 2001) and single-cell recording studies with non-human primates (Apicella, Scarnati, Ljungberg, & Schultz, 1992; Hollerman, Tremblay, & Schultz, 1998) have shown that the striatum is active not only when a reward is actually obtained, but also when a reward is simply anticipated. Moreover, as mentioned above, we have previously shown that the striatum is recruited by both the materialistic reward of money and the abstract social reward of a good reputation (Izuma et al., 2008).

In order to test the idea that during reputation processing the mPFC plays an essential role in representing how others view us (reputation) while the value associated with the reputation is processed in the striatum, 26 subjects underwent fMRI scanning while in a situation where there was a strong demand for subjects to process their own reputation. In the present study, subjects were asked to perform three experimental tasks, which were similar to the tasks used in previous neuroimaging studies investigating the neural basis of self-reference (D'Argembeau et al., 2007; Fossati et al., 2003; Kelley et al., 2002; Ochsner et al., 2005; Schmitz et al., 2004). However, unlike previous studies using self-reference tasks with trait adjectives, subjects in the present study were presented with sentence stimuli depicting prosocial or antisocial behaviors that were designed to induce a high concern for their reputation when being observed by others (e.g., "I never drop litter on the street"; for more examples, see Table 1). In this situation, how subjects answer these questions has considerable impact on their reputation. More specifically, subjects performed three types of

TABLE 1
Sample sentences

1. I never hesitate to go out of my way to help someone in trouble (SDS)
2. I can remember "playing sick" to get out of something (SDS)
3. I never cover up my mistakes (IM)
4. I have said something bad about a friend behind his or her back (IM)
5. I always keep my word to others (Original)
6. I am not punctual for appointments (Original)

Notes: The first, third, and fifth sentences depict prosocial behaviors; the second, fourth, and sixth sentences depict antisocial behaviors. All the sentences were translated into Japanese for the experiment. SDS, Social Desirability Scale; IM, Impression Management scale.

task: the ‘‘self-descriptiveness judgment task’’ (‘‘Self’’), in which the subjects were asked to rate the extent to which each sentence described them, and thus show their behavioral tendencies relative to social norms; the ‘‘social-appropriateness judgment task’’ (‘‘Social’’), in which subjects were asked to rate the extent to which a depicted behavior was regarded as socially acceptable, and thus show their understanding of social norms; and the ‘‘letter-search task’’ (‘‘Letter’’), in which the subjects were asked to simply count the number of times a certain Japanese syllabary character or *hiragana* appeared in a sentence. Furthermore, during these tasks the presence of observers was also manipulated in the similar manner to the previous study (Izuma, Saito, & Sadato, in press) in order to induce a strong reputational concern. In half of the four experimental sessions, the subjects were presented with images of observers on the screen, who the subjects believed were sitting in the room next to the fMRI scanner and observing their performance through a video-camera (Figure 1). While in the previous study (Izuma et al., in press) subjects performed only one task

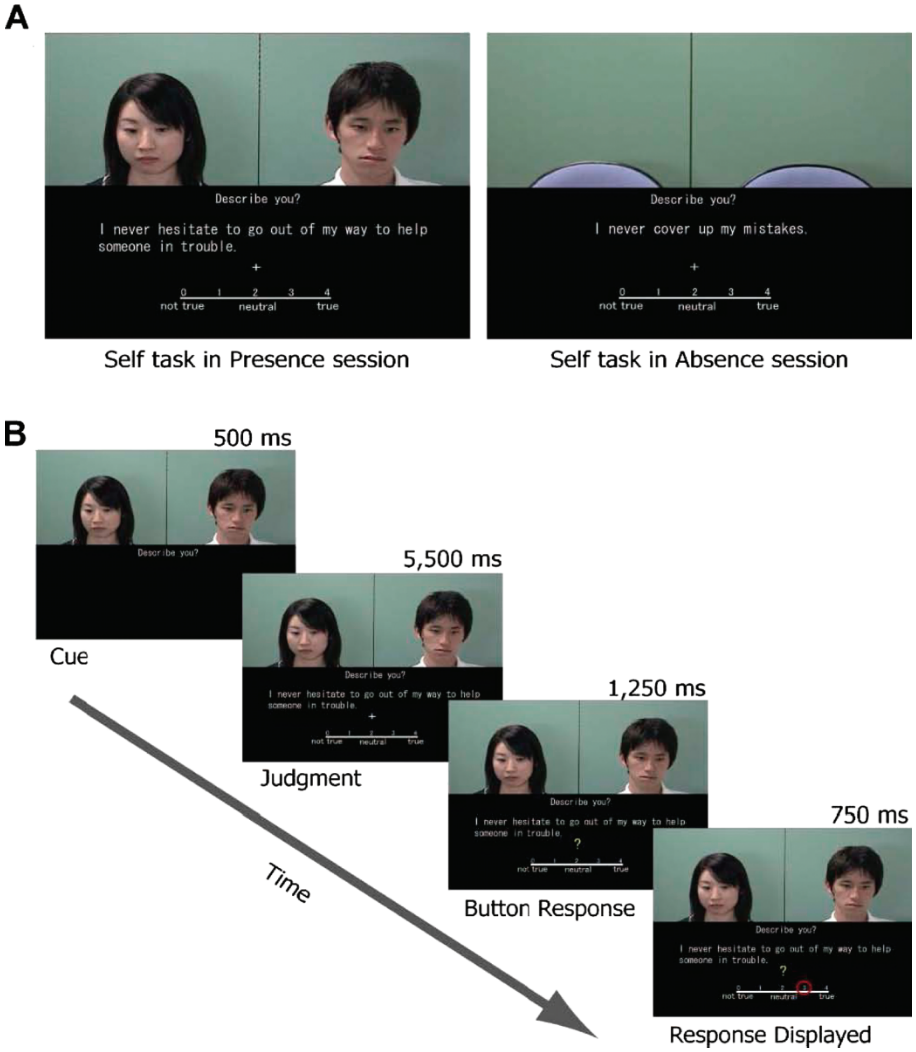
A
Describe you?
I never hesitate to go out of my way to help someone in trouble.
+
0 1 2 3 4
not true neutral true
Describe you?
I never cover up my mistakes.
+
0 1 2 3 4
not true neutral true
Self task in Presence session
Self task in Absence session
B
500 ms
Describe you?
Cue
5,500 ms
Describe you?
I never hesitate to go out of my way to help someone in trouble.
+
0 1 2 3 4
not true neutral true
Judgment
1,250 ms
Describe you?
I never hesitate to go out of my way to help someone in trouble.
?
0 1 2 3 4
not true neutral true
Button Response
750 ms
Describe you?
I never hesitate to go out of my way to help someone in trouble.
?
0 1 2 3 4
not true neutral true
Response Displayed
Time

Figure 1. Experimental stimuli. (A) A single frame of the judgment period of one Self trial during the Presence session (left). The top half of the screen shows a video image of two other people (actors), who the subjects believed were watching their performance in the room next to the fMRI scanner. The bottom half of the screen shows an instructional cue, a sentence, a cross hair, and the five-point scale for the Self trial. Similarly, a single frame of the judgment period of one Self trial during the Absence session is shown on the right. A video showing the upper portions of two chairs, instead of the two actors, was played during the Absence sessions. (B) Sequence of one self-descriptiveness judgment trial in the Presence session (8 s). Importantly, the subject's response to each sentence was displayed on the screen (a red circle) so that it was clearly observable by others. Each subject completed three trials (responding to three different sentences) in one Self block (24 s). For the Social and Letter trials within the same session, the sequence and stimuli were identical except for the instructional cue and scale.

(the donation task) during the presence or absence of observers, here subjects were asked to perform three different tasks upon seeing the same sentence stimuli. This 2 (observer; presence or absence) × 3 (task; Self, Social, or Letter) factorial design made it possible to investigate the effect of the presence of others on the brain activity during a specific task, while ruling out simple arousal effects due to the presence of observers.

In the present experimental paradigm, we predicted that if the mPFC and the striatum play essential roles in processing one's own reputation, we should observe coactivation of the mPFC and the striatum when subjects disclose their social attitudes (Self and Social tasks) in front of other people. As mentioned above, the role of the mPFC in the formation of metarepresentations has been implicated in self-referential processing (D'Argembeau et al., 2007; Fossati et al., 2003; Johnson et al., 2002; Kelley et al., 2002; Ochsner et al., 2005; Schmitz et al., 2004; Zysset et al., 2002), and in the current experiment, the Self tasks are thought to be inherently more directly related to one's reputation than their responses on the Social task. Therefore, we expected stronger mPFC activation during the Self task compared to the Social task (and the Letter task) regardless of the presence of observers, and anticipated that the area of the mPFC involved in metacognitive representation would be included in the region activated by this Self vs. Social contrast. More critically, we expected that within the mPFC area activated by the Self vs. Social contrast, there would be areas whose activity during the Self task is enhanced when observers are present. In addition to self-referential processing (i.e., subjects' views of themselves), the presence of observers requires subjects to process how they are seen by others (i.e., reflected self-knowledge) and how they would like others to view them. Thus, we predicted that if the mPFC's function of forming metacognitive representation plays a crucial role in processing one's own reputation, it should show stronger activation when the social situation increased the demand on processing one's own reputation; that is, when subjects perform a task (especially the Self task) in front of others in which how they respond to questions greatly impacts their reputation. On the other hand, the Letter condition offers an ideal control condition to examine the effect of observers, because observers should have little influence on subjects' reputational concern during this task. In the Self and Social tasks, subjects disclose their behavioral tendency relative to social norms, or express their understanding of social norms, both of which should affect their reputation (for example, we are likely to distrust someone who says that "playing sick" is socially acceptable or something that he or she often does). In contrast, in the Letter task the correct answer is objectively defined, and how subjects perform the task does not reflect their personality (such as trustworthiness) or social knowledge. Also, as humans possess a strong drive to seek a good reputation, which is easily enhanced by the presence of observers (Bateson et al., 2006; Haley & Fessler, 2005; Kurzban et al., 2007), we expected higher activations in the striatum when subjects anticipated the reward of a positive reputation when presenting themselves in front of others.

MATERIALS AND METHODS

Participants

A total of 28 healthy right-handed naïve subjects participated in the fMRI study. The reported analyses were based on 26 subjects (16 males and 10 females; mean age 24.1 ± 3.4 years). Two subjects were excluded from the analysis due to excessive head motion. None of the subjects had a history of neurological or psychiatric illness. All of the subjects gave written informed consent for participation, and the study was approved by the Ethical Committee of the National Institute for Physiological Sciences (NIPS), Japan.

Procedure for the fMRI study

When the subjects arrived at the scanner control room, they were informed that they would take part in various judgment tasks within the fMRI scanner. Furthermore, they were advised that two other people (actors, one male and one female, who were around the same age as the subjects) were participating in the study with them, and that in two of the four sessions these two people would monitor their performance through a video camera in the room next to the fMRI scanner. Outside the scanning room, all of the subjects were shown the experimental setup, which consisted of two chairs, a video camera, a screen, and a screen splitter. The subjects were told that their responses during scanning would be shown on the screen, and that they would see the faces of the two actors who were sitting in the chairs and watching their performance during the task (Figure 1A, left). The subjects were also told that when these actors were on the screen, they were engaged in an ‘‘impression formation task’’. The details of this task were not explained to the subjects, but it was emphasized that the two actors would be paying attention to their performance. Also, the subjects were told that when the actors were not on the screen, and only the upper portions of the two chairs were visible (Figure 1A, right), the observers would be engaged in different tasks that were unrelated to the subject's task.

After completing an 8-min practice session in the scanner control room (the sentences used in the practice session were not included in the main paradigm), all of the subjects met and were introduced to at least one of the two actors before entering the fMRI room. If one of the actors could not meet a subject, that individual was told that the actor was on his or her way and would arrive shortly. Moreover, in order to make the interaction with the two actors during scanning more meaningful, the subjects were told that they would engage in some other tasks together with the two actors after the fMRI scanning. All of the subjects believed that the two actors were in the room next to the fMRI scanner room during the task (this was confirmed during an interview at the end of the experiment); however, in reality, a pre-recorded video was used throughout the sessions. In addition, in order to control for possible confounding factors elicited by changes in the actors' facial expressions or head and eye movements, their expressions were kept neutral, and their gazes were fixed on the bottom half of the screen, as if they were looking at the subjects' responses. Head and eye movements were kept to a minimum throughout the video (all subjects were told in advance that the two actors would be asked to stay still and to focus on their task, in order not to distract the subjects). After scanning, all of the subjects were fully debriefed as to the purpose of the study and the use of the video.

Experimental paradigm

A 2 × 3 factorial design was used in this study. The first factor was whether the subjects performed the tasks in the presence of actors observing their performance (Presence) or in their absence (Absence). The second factor was the kind of task in which the subjects were engaged: Self, Social, or Letter.

All of the subjects completed a total of four sessions. The video of the two actors was shown on the top half of the screen throughout the two Presence sessions (Figure 1A, left). The video showing the upper portions of the chairs was shown throughout the two Absence sessions, so it was clear to the subjects that nobody was watching their responses (Figure 1A, right). It should be noted that because the video of observers or chairs was constantly played throughout a session, there was no common baseline condition (such as a screen containing only a fixation cross) between the Presence and Absence sessions. Since it is highly likely that subjects engaged in mentalizing or inferring the mental states of two observers during the rest block of the Presence sessions (two observers and a fixation cross were presented), activations during the Self and Social task were plotted relative to the Letter task of the same sessions (Presence or Absence), and not the fixation condition, in order to make comparisons between the Presence and Absence sessions more valid (Figures 2B, 3B).

During each session, the subjects performed the tasks shown on the bottom half of the screen. They completed three types of task block (Self, Social, and Letter), each of which comprised three sentences (trials), and a Rest block. In total, 60 sentence stimuli were used, 33 of which were drawn from the Social Desirability Scale (SDS) (Crowne & Marlowe, 1960), 17 of which were drawn from the Impression Management (IM) scale (a subscale of the Balanced Inventory of Desirable Responding measure) (Paulhus, 1984), and 10 of which were original. Each sentence depicted either rarely or commonly occurring culturally approved behaviors (see Table 1). Some sentences from the SDS and IM scales were modified so that half of the sentences depicted prosocial behaviors, and the other half depicted antisocial behaviors. The prosocial and antisocial sentences were distributed equally, and 30 different sentences were used across the Presence and Absence sessions. However, within each subject, the 30 sentences used in the Presence and Absence sessions were the same across the three experimental tasks (Self, Social, and Letter).

In the Self blocks, the subjects were instructed to rate the extent to which a sentence described them using a five-point scale (in which 0 was ‘‘not true’’, 2 was ‘‘neutral’’, and 4 was ‘‘true’’), by

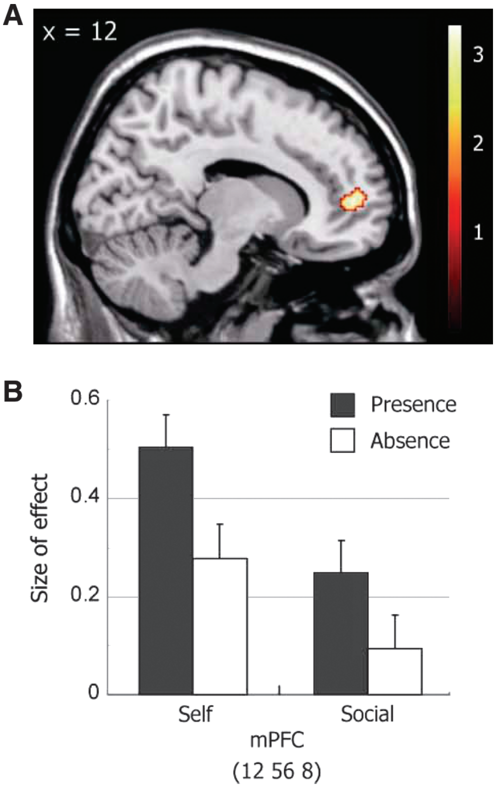

Figure 2. Group activation results of the observer effect in the mPFC. (A) Within *a priori* ROIs in the mPFC, significant activation was found by the interaction contrast of (Presence – Absence) × (Self – Letter). The statistical threshold within the ROIs was set at $p < .005$ (uncorrected) and cluster $p < .05$ (corrected for multiple comparisons). The scale shows t values. (B) Bar graphs indicate the effect sizes in the mPFC during the Self and Social conditions in the Presence or Absence sessions relative to the corresponding Letter condition of the Presence or Absence sessions. Error bars indicate the standard error of the mean (SEM).

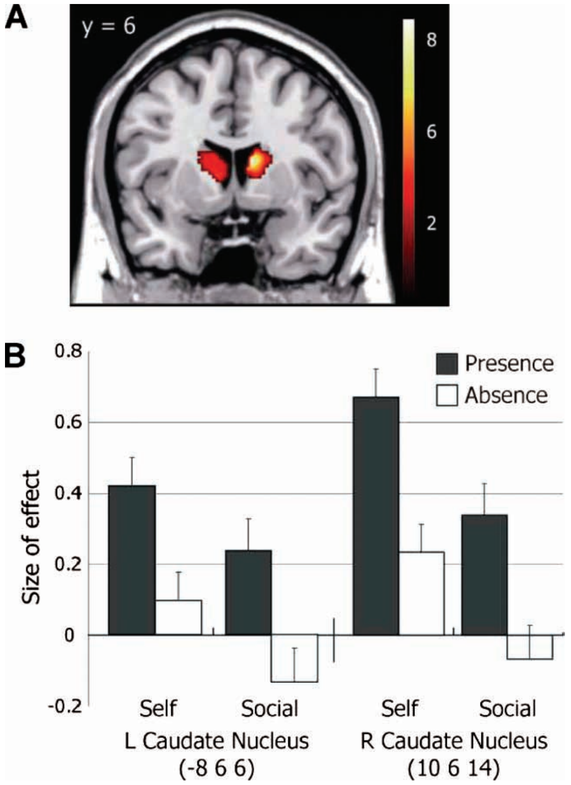

Figure 3. Group activation results of the observer effect in the striatum. (A) The head of the caudate nucleus bilaterally was significantly activated by the interaction contrast of (Presence – Absence) × (Self – Letter). Within the striatal ROIs, the statistical threshold was set at $p < .005$ (uncorrected) and cluster $p < .05$ (corrected for multiple comparisons). The scale shows t values. (B) Bar graphs indicate the effect sizes in bilateral caudate during the Self and Social conditions in the Presence or Absence sessions relative to the corresponding Letter condition of the Presence or Absence sessions. Error bars indicate the standard error of the mean (SEM).

making a button press with their right hand. In the Social blocks, the subjects rated the extent to which a behavior depicted in a sentence was regarded as socially acceptable using a five-point scale (in which 0 was "wrong", 2 was "neutral", and 4 was "right"). In the Letter blocks, the subjects were instructed to count the number of a certain Japanese syllabary characters or *hiragana* in a sentence using a five-point scale (in which 0 was "none", 2 was "two", and 4 was "more than three"). Because the sentence stimuli in the present study were written in Japanese, and included not only *hiragana* characters but also Chinese characters or *kanji*, the subjects had to silently read the sentence in order to perform the Letter task successfully. Therefore, the Letter task was analogous to a task in which subjects search for a certain sound or phoneme (rather than a certain letter of the alphabet) in an English sentence. In the Rest blocks, the subjects were instructed to simply look at the fixation cross on the bottom half of the screen, and not to press any buttons.

For the Self, Social, and Letter blocks, each of the three trials in each block began with an instructional cue indicating which kind of task the subject should perform. The instructional cue remained on the screen for the rest of the block. After 500 ms, a sentence, a cross-hair, and the five-point scale for the task were shown on the bottom half of the screen for 5,500 ms. Then, the cross-hair was replaced with the symbol "?" for 1,250 ms, within which time period the subjects were instructed to respond with their right hand using a five-button response box. After that, a red circle was displayed around the number that the subject had just chosen for 750 ms (Figure 1B; if the subjects did not respond within the time limit, the number that they had chosen on the previous trial was circled again). In a practice session before scanning, the subjects were instructed to make a decision within the judgment period of 5,500 ms, and to respond as quickly as possible when the fixation cross was replaced by the symbol "?", in order to avoid an inappropriate response display. After every three blocks (Self, Social, and Letter blocks) there was a Rest block, in which the fixation cross was displayed in the center of the bottom half of the screen for 24 s. Every session comprised 8-min scans, each consisting of pseudorandomly ordered blocks (four conditions in a session, five blocks per condition, three trials per block, and 8 s per trial), and three different orders of pseudorandomized blocks were used across subjects. One-half of the subjects completed two Presence sessions followed by two Absence sessions; the other half of the subjects performed the experiment in the reverse order.

All of the stimuli for the tasks were prepared and presented using Presentation software (Neurobehavioral Systems, CA) on a microcomputer (Dimension 8200, Dell Computer Co., TX). The videos showing the two actors and the two chairs were played by a digital video-cassette player (GV-D1000, Sony, Tokyo, Japan). Using a liquid crystal display (LCD) projector (DLA-M200L, Victor, Yokohama, Japan), the visual stimuli were projected onto a half-transparent viewing screen via a screen splitter (MV-40F, FOR-A, Tokyo, Japan) so that the video was shown on the top half of the screen, and the stimuli for the tasks were presented on the bottom half of the screen. The screen was located behind the head coil, and the subjects viewed the stimuli through a mirror. All of the stimuli for the tasks were originally written in Japanese and presented as white letters against a black background.

Image acquisition

Images were acquired using a 3 T MR imager (Allegra, Siemens, Erlangen, Germany). Functional images were acquired using interleaved T2*-weighted gradient-echo echo-planar imaging (EPI) sequences to produce 44 continuous 3-mm thick trans-axial slices covering the entire cerebrum and cerebellum (repetition time [TR] = 3,000 ms; echo time [TE] = 25 ms; flip angle [FA] = 85°; field of view [FOV] = 192 mm; 64 × 64 matrix; voxel dimensions = 3.0 × 3.0 × 3.0 mm). A high-resolution anatomical T1-weighted image was also acquired for each subject with magnetization-prepared rapid gradient-echo (MPRAGE) imaging (TR = 2.5 s; TE = 4.38 ms; FA = 8°; 256 × 256 matrix; 192 slices; voxel dimensions = 0.75 × 0.75 × 1 mm).

Imaging data analysis

After discarding the first four volumes to allow for stabilization of the magnetization, the remaining 160 volumes per session (a total of 640 volumes per subject for four sessions) were used for analysis. The data were analyzed using Statistical Parametric Mapping 5 (SPM5, Wellcome Department of Imaging Neuroscience, London, UK) (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007) software implemented in Matlab 7.1 (Mathworks, Sherborn, MA). Head motion was corrected for using the realignment program of SPM5 (Friston et al., 1995). Following realignment, the volumes were normalized to Montreal Neurological Institute (MNI) space using a transformation matrix obtained from the normalization process of the first EPI image of each individual subject to the EPI template. The normalized fMRI data were spatially smoothed with a Gaussian kernel of 8 mm (full width at half maximum) in the *x*, *y*, and *z* axes.

Statistical analysis was conducted at two levels. First, individual task-related activation was evaluated. Second, the summary data for each individual was incorporated into a second-level analysis using a random-effects model (Friston, Holmes, & Worsley, 1999) to make inferences at a population level.

In the individual analyses, the signal was scaled proportionally by setting the whole-brain mean value to 100 arbitrary units. The signal time course for each subject was modeled with a general linear model. Six regressors of interest (condition effects; 2 [presence/absence of observers] × 3 [experimental tasks; Self, Social, or Letter]) were generated using a box-car function convolved with a hemodynamic-response function. Regressors that were not of interest, such as the session effect, and high-pass filtering (128 s) were also included. To test hypotheses about regionally specific condition effects, the estimates for each condition were compared by means of the linear contrasts shown in Table 2.

The weighted sum of the parameters estimated in the individual analyses consisted of ‘‘contrast’’ images, which were used for the group analyses with a random-effects model. The contrast images obtained by each individual analysis represented the normalized increment of the fMRI signal for each subject. The SPM{t} for the contrast images was created as described above. Significant signal changes for each contrast were assessed by means of *t*-statistics on a voxel-by-voxel basis.

Since our hypothesis focused on the role played by the mPFC and striatum in reputation processing, we performed region of interest (ROI) analysis. For the mPFC, the ROIs were defined both anatomically and functionally. We first generated the anatomical ROIs (anterior cingulate and superior medial frontal cortex) using the WFU PickAtlas toolbox for SPM (Maldjian, Laurienti, Kraft, & Burdette, 2003) with a dilation factor of 1. Then, in order to functionally define the ROIs in the mPFC, the anatomical ROIs were intersected with the voxels showing significant activation ($p < .001$, uncorrected for multiple comparisons) in the Self vs. Social contrast. Since the mPFC has been implicated in the formation of metacognitive representations in the self-referential task (Amodio & Frith, 2006), our mPFC ROIs included the particular area of the mPFC involved in metacognitive representation. For the striatum, we anatomically defined the ROIs (caudate and putamen) using the WFU PickAtlas toolbox. Within these ROIs, the statistical threshold was set at $p < .005$ (uncorrected) and a cluster $p < .05$ (corrected for multiple comparisons). For descriptive purposes, the areas activated by the main effects of task and observers were reported at a threshold of $p < .001$ (uncorrected) with an extent threshold of more than 30 contiguous voxels. Activations in the interaction contrast outside of the ROIs were also reported if they exceeded the same threshold ($p < .001$ and $k > 30$ voxels).

RESULTS

Behavioral results

The mean ± standard deviation (*SD*) percentage of failed trials was 3.65 ± 2.2%, indicating that the subjects responded within the time limit of 1,250 ms and made appropriate responses on ~96% of the trials. A 2 (Presence/Absence of observers) × 3 (experimental tasks; Self, Social, or Letter) repeated-measures analysis of variance (ANOVA) on the percentage of failed trials revealed a significant main effect of task alone, $F(2, 50) = 4.03$, $p = .024$, and the *post hoc* Bonferroni comparisons revealed that subjects were significantly more likely to fail to respond within the time limit on the Social task (mean ± $SD = 4.68 \pm 3.97\%$) compared to the Self task (mean ± $SD = 2.56 \pm 2.12\%$, $p = .007$). The results of the Letter task fell somewhere between the two, with a mean ± *SD* of 3.72 ± 2.92%. Neither the main effect of observers nor the interaction effect was significant (both *p* values > .42, n.s.), indicating that the presence/absence of observers did not affect the number of failed trials.

The result of the 2 (Presence/Absence of observers) × 3 (experimental tasks; Self, Social, or Letter) repeated-measures ANOVA on reaction time data revealed a significant main effect of task, $F(2, 50) = 10.2$, $p < .001$. The *post hoc*

TABLE 2
Predefined contrasts

	Conditions					
	Presence			*Absence*		
Contrast	*Self*	*Social*	*Letter*	*Self*	*Social*	*Letter*
Self – Social	1	−1	0	1	−1	0
Self – Letter	1	0	−1	1	0	−1
Pre – Abs	1	1	1	−1	−1	−1
(Pre – Abs) × (Self – Letter)	1	0	−1	−1	0	1
(Abs – Pre) × (Self – Letter)	−1	0	1	1	0	−1
(Pre – Abs) × (Self – Social)	1	−1	0	−1	1	0
(Pre – Abs) × (Self + Social – 2Letter)	1	1	−2	−1	−1	2

Notes: Pre, the Presence condition; Abs, the Absence condition.

Bonferroni comparisons showed that subjects were significantly faster to respond in the Letter task (mean ± *SD* = 475.2 ± 74.6 ms) compared to both the Self task (mean ± *SD* = 513.1 ± 88.0 ms, $p = .002$) and the Social task (mean ± *SD* = 508.6 ± 85.2 ms, $p = .001$), but the difference between the Self and Social tasks was not significant ($p = 1.0$, n.s.).

The mean ± *SD* percentage accuracy for the Letter search task was 83.5 ± 7.1%, and there was no significant difference in task performance between the Presence (82.9 ± 8.07%) and Absence (84.0 ± 8.00%) sessions ($p = .50$, n.s.).

The means (± *SD*) of the social-appropriateness judgment ratings for the Presence and Absence sessions were 4.36 (± 0.38) and 4.37 (± 0.27) respectively (subjects' responses to antisocial behaviors were reverse-scored so that a higher score indicates a stronger tendency to choose "right" for prosocial items and "wrong" for antisocial items). In both conditions, these scores were significantly higher than the midpoint of 3 on the 5-point scale (both *p* values < .001), and the scores were close to the maximum value of 5, suggesting that the subjects had a clear understanding of social norms. Also, the means (± *SD*) of the self-descriptiveness judgment ratings for the Presence and Absence sessions were 3.48 (± 0.46) and 3.55 (± 0.48) respectively (subjects' responses to antisocial behaviors were reverse-scored so that a higher score indicates a stronger tendency to choose "true" for prosocial items and "not true" for antisocial items). These scores were both significantly higher than the midpoint value (both *p* values < .001), suggesting that the subjects responded to items in a socially desirable manner.

fMRI results

Before investigating the effect of observers, the main effect of the different tasks was explored in order to see if we successfully replicated the previous finding as to the neural basis of self-referential processing. Consistent with previous studies (D'Argembeau et al., 2007; Fossati et al., 2003; Johnson et al., 2002; Kelley et al., 2002; Ochsner et al., 2005; Schmitz et al., 2004; Zysset et al., 2002), the contrast of the Self vs. Social (and Self vs. Letter) conditions revealed significant activations in the mPFC and posterior cingulate cortex (Figure 4). Also, when the main effect of observers (Presence vs. Absence) was explored using the fixation rest block as an implicit baseline, activations were found in the lingual gyrus ($x = -4$, $y = -76$, $z = -6$) and middle frontal gyrus ($x = 32$, $y = 44$, $z = -2$) (figure not shown). The activation in the primary visual cortex can be explained by the presence of more visual stimuli (faces of observers) in the peripheral visual field when the subjects engaged in tasks in the Presence condition compared to the Absence condition. Because the focus of the present study is on the effect of observers on a particular task, these results are not discussed further.

We investigated the effect of observers during the Self task with the Letter task as a control within the *a priori* ROIs in the mPFC, which consisted of a total volume of about 4,200 voxels. The interaction contrast of (Presence – Absence) × (Self – Letter) revealed significant activations, as predicted ($x = 12$, $y = 56$, $z = 8$, 179 voxels, Figure 2A). To determine whether the presence of observers affected the Self and Social tasks differently, we compared the effect sizes at the mPFC peak between the Presence and Absence sessions for both the Self and Social conditions relative to the corresponding Letter condition sessions (Figure 2B). A 2 (observer; Presence vs. Absence) × 2 (task; Self vs. Social) repeated-measures ANOVA revealed highly significant main effects of both task, $F(1, 25) = 22.1$, $p < .001$, and observer, $F(1, 25) = 13.2$, $p = .001$, but no significant interaction, $F(1, 25) = 0.93$, $p = .34$, n.s. We then explored the same interaction contrast of (Presence – Absence) × (Self – Letter) within the striatum ROIs, which consisted of a total volume of about 5,400 voxels. This also revealed significant activations in the head of the caudate nucleus bilaterally (right caudate = 225 voxels, left caudate = 204 voxels, Figure 3A). We then plotted the effect sizes at the peaks in the caudate nucleus of the Presence and Absence sessions for both the Self and Social conditions relative to the corresponding Letter condition sessions. These findings showed similar activation patterns as the mPFC (Figure 3B). A 2 (observer; Presence vs. Absence) × 2 (task; Self vs. Social) repeated-measures ANOVA revealed that both the right and left caudate nucleus showed highly significant main effects of both task and observer (all *p* values < .002), but no significant interactions were found in either area (both *p* values > .71, n.s.). It should be noted that activations in the mPFC and the bilateral caudate nucleus showed no gender difference. When we included subjects' gender as another factor, there were no significant two-way interactions of gender by task and gender by observer (all p values > .16) nor three-way interactions (all p values > .56), suggesting the experimental tasks and the presence of observers similarly affected mPFC and striatal activations in both genders. Outside the ROIs, the (Presence – Absence) × (Self – Letter) contrast revealed activations in anterior cingulate cortex ($x=6$, $y=20$, $z=18$), middle frontal gyrus ($x=36$, $y=20$, $z=28$), hypothalamus ($x=2$, $y=-4$, $z=-16$), and cerebellum ($x=-4$, $y=-36$, $z=-20$). No regions showed significant activations in the reverse interaction contrast of (Absence – Presence) × (Self – Letter).

The interaction contrast of (Presence – Absence) × (Self – Social) was investigated, and we found no significant activations within or outside the mPFC and striatum ROIs (activation was found only in the right cerebellum when the threshold was changed to $p < .01$ [uncorrected]). Since the Self and Social tasks showed similar activation patterns in terms of the effect of observers, we also explored the contrast of (Presence – Absence) × (Self + Social – 2Letter) and found that besides the abovementioned activations in mPFC and bilateral caudate nucleus, right amygdala ($x=18$, $y=2$, $z=-16$, 60 voxels) and right anterior insula ($x=32$, $y=20$, $z=-8$, 84 voxels) were also activated, which seems to reflect the processing of negative emotion (e.g., fear and anxiety) associated with being evaluated by others. Other activated areas in this contrast include right middle frontal gyrus, bilateral middle temporal gyrus, left fusiform gyrus, and cerebellum.

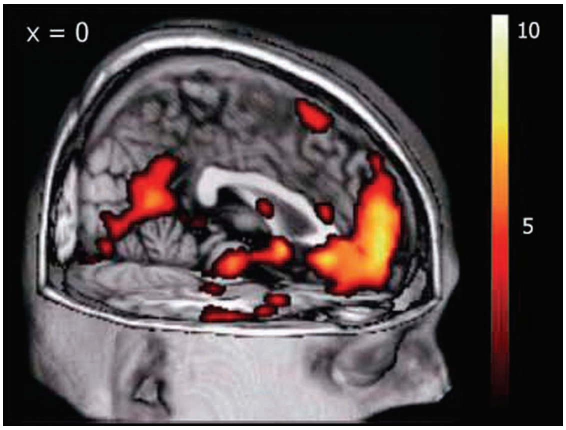

Figure 4. Brain regions activated by the (Self – Social) contrast. This contrast was inclusively masked by the (Self – Letter) contrast. mPFC and posterior cingulate regions were widely activated. The statistical threshold was set at $p < .001$ (uncorrected) for height and $k > 30$ voxels. The scales show t values.

DISCUSSION

In the present study, we investigated the roles of the mPFC and striatum in processing one's own reputation by having subjects perform tasks in the presence or absence of observers. First, our results replicated previous studies by showing that the mPFC was widely activated during the Self task compared to the Social and Letter tasks. Second, the present study provides the first evidence of enhanced mPFC activity by the mere presence of observers during the Self and Social tasks relative to the Letter task. In addition to the mPFC, the dorsal striatum (caudate nucleus) showed similar activation patterns; the presence of observers led to a highly significant increase in activity during the Self task (and the Social task). The coactivation found in the mPFC and the striatum in the social situation, where there was a strong demand for subjects to process how they are viewed by other people, supports the idea that these two brain areas play a crucial role in reputation processing.

The present result can be seen as direct evidence for the idea that the mPFC plays a key role in the processing of one's own reputation (Amodio & Frith, 2006; Frith & Frith, 2008). While involvement of the mPFC in reflected self-knowledge, the process of seeing oneself from another person's perspective, has previously been reported (D'Argembeau et al., 2007; Ochsner et al., 2005), these studies simply manipulated the instructions given to subjects (e.g., explicitly asking them to judge the extent to which a friend would perceive that an adjective described them). In contrast, our current study manipulated the situational factor of the presence/absence of observers, which is known to greatly influence an individual's concern for their reputation (Bateson et al., 2006; Haley & Fessler, 2005; Kurzban et al., 2007). Subjects in the present study were never instructed to consider observers' mind or view themselves from the perspective of observers, and they were told to ignore the observers and concentrate on their own tasks (Self, Social, or Letter). Thus, the present study revealed that the mPFC is automatically activated when there is a greater demand for processing one's own reputation. This result is consistent with the idea that the mPFC contributes to the formation of metacognitive representations that are crucial to the representation of one's reputation. Also, previous findings on the neural correlates of reflected self-knowledge are not sufficient to explain the neural processes underlying reputation processing, because these studies did not consider the importance of the associated reward value (see below).

Although being watched by other people is very natural for humans, and its effect on behaviors has been widely acknowledged by social scientists, to our knowledge only one previous study manipulated the presence of observers while measuring neural activity, and this study also indicated mPFC involvement when individuals worried about how they were viewed by others (Amodio, Kubota, Harmon-Jones, & Devine, 2006). In this event-related potential (ERP) study, subjects engaged in a stereotype inhibition task to measure the level of racial bias in private or in public (Amodio et al., 2006). In the private condition, subjects were told that their responses would remain confidential, while in the public condition they were told that the experimenter would pay attention to their performance and check to determine if they showed signs of racial prejudice. This study was very similar to our present study in that subjects performed tasks in which how they responded strongly influenced their social image, and the degree to which their responses were viewed by other people was manipulated. Amodio et al. found that a larger error-related negative component, linked to the dorsal anterior cingulate cortex (dACC or dorsal part of mPFC), predicted better response control in both conditions, whereas a larger error-related positive component, linked to the rostral ACC (rACC; presumably the same region as the mPFC in the present study), predicted better response control only in the public condition. The latter effect was seen only among subjects who cared about their social image and tried to appear non-prejudiced (Amodio et al., 2006). Taken together, the results of their ERP study and our fMRI data converge to suggest that the mPFC (or the anterior rostral part of the ACC) is involved in the formation of a complex and abstract representation which is needed to process how one is viewed by others (Amodio & Frith, 2006).

Furthermore, the similar activation patterns found in the mPFC and the striatum support the idea that both the representation of one's own reputation in the mPFC and the valuation process in the striatum are crucial to social reward (reputation) processing (Izuma et al., 2008). It is because of this reputation-related reward value that reputational concern affects human social behaviors (Benabou & Tirole, 2006). As mentioned above, previous behavioral studies indicated that even pictures of eyes or eye-like stimuli were sufficient to increase individuals' drives to seek good reputations (Bateson et al., 2006; Haley & Fessler, 2005). Thus, it was likely that the expectation of a good reputation or the avoidance of a bad reputation were very strong in the present experiment, because subjects were faced with two real observers. In the present study, the activities in the striatum during the Self (and Social) task were enhanced by the presence of observers, which is consistent with our prediction that the expected reward value of a good reputation is represented in the striatum. Therefore, the striatal activations found in the present study are consistent with a previous report showing that the striatum (caudate nucleus) is activated when individuals expected to obtain monetary reward and avoid monetary loss (Knutson et al., 2001; Knutson, Westdorp, Kaiser, & Hommer, 2000). Furthermore, we also found that the same dorsal striatal areas showed higher activation during the Self task compared to the Social task, although the striatum is not typically activated by self-referential processing. However, because even subtle manipulations of anonymity impact the concern for one's reputation (Bateson et al., 2006; Crowne & Marlowe, 1960; Haley & Fessler, 2005; Kurzban et al., 2007; Lautenschlager & Flaherty, 1990; Nederhof, 1985; Paulhus, 1984), it is extremely difficult or even impossible to completely exclude this concern from participants in any fMRI study, as subjects are likely to believe that their performance is monitored at least by the experimenter, even in the absence of explicit observers. Thus, it is plausible that the expectation of social reward was higher during the Self task, in which the subjects directly showed their social attitudes, than during the Social task, in which subjects only indicated their understanding of social norms, regardless of whether they follow such norms.

It should be noted, however, that although it is tempting to conclude that the strong striatal activation during self-presentation (the Self and Social tasks) in front of others reflects not just the expectation of good reputation but the strong motivation to present themselves in a positive manner, this interpretation is not valid because we did not find behavioral evidence to support it in the current study (there is no difference in ratings of self-descriptiveness judgment between the Presence and Absence conditions). However, the main reason for the null result seems to be the large variance created by the different items used between two conditions. In the previous study where the same observer manipulation was used (Izuma et al., in press), subjects were presented with the same charities during the Presence and Absence conditions, and we found that they actually donated more often in the Presence condition than the Absence condition suggesting heightened motivation for social reward in the Presence condition.

As the striatum is known to play a central role in value-based decision-making (Samejima, Ueda, Doya, & Kimura, 2005; Tom, Fox, Trepel, & Poldrack, 2007), and the mPFC plays a role in metacognitive representations, which allow us to reflect on the values associated with outcomes and actions in social situations (Amodio & Frith, 2006), we suggest that these two regions of the brain are crucial to guiding behavior in social interactions. We argued that, for complete reputation processing, reputation or how one is viewed by other people would first be represented in the mPFC, and this information is sent to the striatum where its value is further processed in order to select an appropriate action in a given social situation. Consistent with this idea, there is a direct anatomical connection between mPFC and striatum in monkey (Haber, Kunishio, Mizobuchi, & Lynd-Balta, 1995), and similar connectivity was reported in human using probabilistic diffusion tractography (Draganski et al., 2008). The present findings are consistent with the idea that the mentalizing function of the mPFC and reward-related brain areas such as striatum are both important for human social decision making (Lee, 2006; Walter, Abler, Ciaramidaro, & Erk, 2005). However, our present findings particularly stress that these brain areas are critical not only for evaluating others in social interactions (e.g., King-Casas et al., 2005; Rilling et al., 2004), but also for evaluating how we ourselves are evaluated by others, both of which are key pieces of information in deciding how to behave in social interactions.

The present findings also provide valuable insight into the relationship between human mPFC function and the evolution of human cooperation. While the role of the striatum in reward processing is shared by other animals, such as monkeys and rats (Berridge & Robinson, 2003), the function of the mPFC in representing other's thoughts seems to be uniquely human. It is an interesting coincidence that, among animals, only humans are considered to have a theory of mind (see Call & Tomasello, 2008; Heyes, 1998 for further discussion on this issue) and cooperate with genetically unrelated individuals. Theoretical research on the evolution of human cooperation suggests that indirect reciprocity is a major mechanism for the evolution of behavior that benefits others through natural selection, and this process requires organisms to assess the reputation of all possible exchange partners and behave differently depending on their reputation (Nowak, 2006). These researchers further speculate that the selective pressure caused by indirect reciprocity may be what led to the evolution of uniquely human cognition (Nowak, 2006; Nowak & Sigmund, 2005). Thus, the lack of the mPFC's ability to form second-level representations might be another constraint that limits cooperation in non-human animals, along with other cognitive limitations (Stevens & Hauser, 2004). Therefore, this mPFC function, together with the value representation in the striatum, enables one to represent complex and abstract goals such as creating a certain impression of oneself in the eyes of others, and thus might be what makes uniquely human cooperation possible.

In conclusion, by using fMRI to recreate a situation that we all encounter in our daily lives (that is, the presence of someone watching us), our findings advance the understanding of how the brain processes an individual's own reputation in everyday social situations. The present study showed that when there is strong demand to process one's own reputation, there is increased activation in the mPFC and the striatum. These results suggest that the mPFC and the striatum are important brain areas in regulating social behaviors and maintaining a cooperative and orderly human society.

Manuscript received 6 May 2009
Manuscript accepted 22 July 2009
First published online day/month/year

REFERENCES

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

Amodio, D. M., Kubota, J. T., Harmon-Jones, E., & Devine, P. G. (2006). Alternative mechanisms for regulating racial responses according to internal vs.

external cues. *Social Cognitive & Affective Neuroscience*, *1*, 26–36.

Apicella, P., Scarnati, E., Ljungberg, T., & Schultz, W. (1992). Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *Journal of Neurophysiology*, *68*, 945–960.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*, 412–414.

Benabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*, 1652–1678.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neuroscience*, *26*, 507–513.

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.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*, 187–192.

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*, 314–325.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354.

D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Balteau, E., Luxen, A., et al. (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience*, *19*, 935–944.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences*, *1104*, 70–88.

Delgado, M. R., Locke, H. M., Stenger, V. A., & Fiez, J. A. (2003). Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 27–38.

Draganski, B., Kherif, F., Kloppel, S., Cook, P. A., Alexander, D. C., Parker, G. J., et al. (2008). Evidence for segregated and integrative connectivity patterns in the human basal ganglia. *The Journal of Neuroscience*, *28*, 7143–7152.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–791.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128.

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., et al. (2003). In search of the emotional self: An FMRI study using positive and negative emotional words. *American Journal of Psychiatry*, *160*, 1938–1945.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *2*, 165–189.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. N., & Penny, W. D. (Eds.). (2007). *Statistical parametric mapping: The analysis of functional brain images*. London: Academic Press.

Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? *NeuroImage*, *10*, 1–5.

Frith, C. D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron*, *60*, 503–510.

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

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*, 77–83.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*, 11–21.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814–821.

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport*, *6*, 1741–1746.

Haber, S. N., Kunishio, K., Mizobuchi, M., & Lynd-Balta, E. (1995). The orbital and medial prefrontal circuit through the primate basal ganglia. *Journal of Neuroscience*, *15*, 4851–4867.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*, 245–256.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*, 101–114; discussion 115–148.

Hollerman, J. R., Tremblay, L., & Schultz, W. (1998). Influence of reward expectation on behavior-related neuronal activity in primate striatum. *Journal of Neurophysiology*, *80*, 947–963.

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*, 284–294.

Izuma, K., Saito, D. N., & Sadato, N. (in press). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of Cognitive Neuroscience.*

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain*, *125*, 1808–1814.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, *14*, 785–794.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*, 78–83.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience*, *21*, RC159.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, *12*, 20–27.

Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*, 103–109.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*, 75–84.

Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology*, *75*, 310–314.

Lee, D. (2006). Neural basis of quasi-rational decision making. *Current Opinion in Neurobiology*, *16*, 191–198.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, *19*, 1233–1239.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 11832–11835.

Milinski, M., Semmann, D., Bakker, T. C., & Krambeck, H. J. (2001). Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proceedings of the Royal Society of London B*, *268*, 2495–2501.

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature*, *415*, 424–426.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*, 1306–1315.

Nederhof, A. D. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*, 263–280.

Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, *8*, 102–107.

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

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573–577.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D., Kihsltrom, J. F., et al. (2005). The neural correlates of direct and reflected self-knowledge. *NeuroImage*, *28*, 797–814.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598–609.

Rilling, J. K., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron*, *35*, 395–405.

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.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*, 1337–1340.

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive & Affective Neuroscience*, *1*, 229–234.

Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *NeuroImage*, *22*, 941–947.

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex*, *10*, 272–284.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*, 60–65.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Walter, H., Abler, B., Ciaramidaro, A., & Erk, S. (2005). Motivating forces of human actions. Neuroimaging reward and social interaction. *Brain Research Bulletin*, *67*, 368–381.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, *288*, 850–852.

Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *NeuroImage*, *15*(4), 983–991.