Brief Communications

Activity in the Superior Temporal Sulcus Highlights Learning Competence in an Interaction Game

Masahiko Haruno and Mitsuo Kawato
ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan

During behavioral adaptation through interaction with human and nonhuman agents, marked individual differences are seen in both real-life situations and games. However, the underlying neural mechanism is not well understood. We conducted a neuroimaging experiment in which subjects maximized monetary rewards by learning in a prisoner's dilemma game with two computer agents: agent A, a tit-for-tat player who repeats the subject's previous action, and agent B, a simple stochastic cooperator oblivious to the subject's action. Approximately 1/3 of the subjects (group I) learned optimally in relation to both A and B, while another 1/3 (group II) did so only for B. Postexperiment interviews indicated that group I exploited the agent strategies more often than group II. Significant differences in learning-related brain activity between the two groups were only found in the superior temporal sulcus (STS) for both A and B. Furthermore, the learning performance of each group I subject was predictable based on this STS activity, but not in the group II subjects. This differential activity could not be attributed to a behavioral difference since it persisted in relation to agent B for which the two groups behaved similarly. In sharp contrast, the brain structures for reward processing were recruited similarly by both groups. These results suggest that STS provides knowledge of the other agent's strategies for association between action and reward and highlights learning competence during interactive reinforcement learning.

Introduction

Humans are extremely interactive creatures and adapt their behavior based on the characteristics of other agents. During behavioral adaptation through interaction with human (Van Lange, 1999) and nonhuman agents (Fogg, 2003), we see differences among individuals in how they learn to cooperate or compete properly in both real-life situations and games.

Two learning strategies are applicable for making decisions through interaction (Barraclough et al., 2004; Lee, 2008). One is to learn the association between one's own action and reward and to select an action associated with the maximum expected reward (Sutton and Barto, 1998). Recent experimental studies have indicated that this mechanism exists in the brain by demonstrating reward prediction signals in the brain (Schultz et al., 2003; Barraclough et al., 2004; Dorris and Glimcher, 2004; O'Doherty et al., 2004; Haruno and Kawato, 2006; Kennerley et al., 2006; Seo and Lee, 2007). In contrast, another strategy predicts the other agent's behavior based on our own action and context and uses it for reward prediction. Such forward modeling of other agent's behavior has been formalized as "mentalizing" or "theory of mind" (Baron-Cohen, 1997; Frith and Frith, 2003) in neuroscience and social psychology and also plays a key role in behavioral game theory (Camerer, 2003) and motor control (Haruno et al., 2001).

When the external environment or the other agent is simple and unaffected by our own behavior, the former strategy is more efficient than the latter. However, when interaction with the other agent is complex, the ability to predict the other agent's behavior plays a crucial role in reinforcement learning and may account for individual differences.

In this paper, we contrast individuals who predict rewards by fully using the other agent's strategy and those who do so less in a neuroimaging experiment of a prisoner's dilemma game (Axelrod, 1984). Here, the subjects maximized monetary rewards by learning to cooperate or defect. We used computer agents because we are interested in the neural mechanism commonly used in interaction with human and nonhuman agents.

The results described here were concisely presented at the 35th annual meeting of the Society for Neuroscience (Haruno and Kawato, 2005).

Materials and Methods

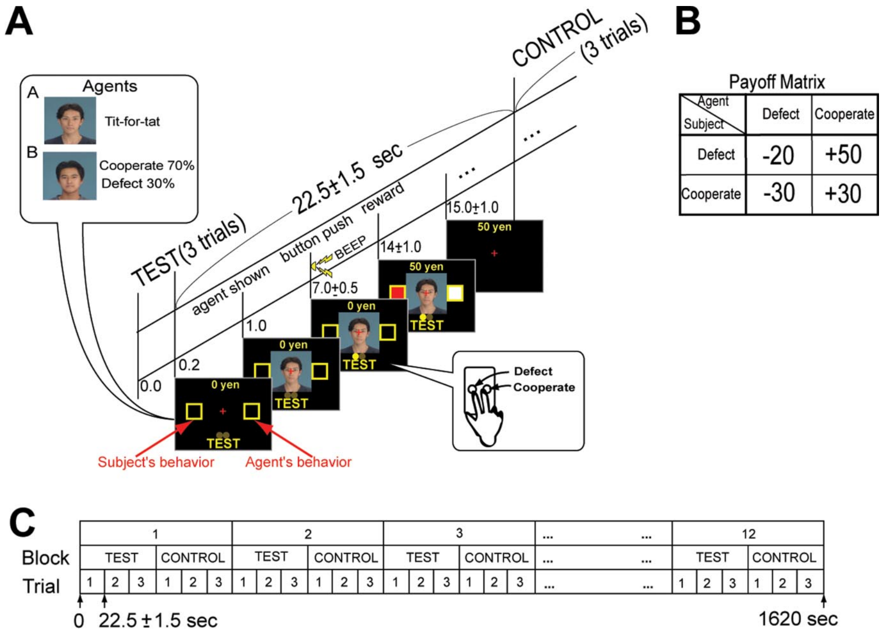
Experimental paradigm. Normal subjects ($n = 32$, 21 males and 11 females, 23–30 years old, and mean ± SD: 25.3 ± 2.51 years) played a prisoner's dilemma game (Axelrod, 1984) against nonhuman computer agents in a functional magnetic resonance imaging (fMRI) scanner. We used two types of agents: agent A, a tit-for-tat player, and agent B, a simple stochastic cooperator with a 0.7 probability of cooperation whose actions were unaffected by the subject's behavior. A's tit-for-tat action only depended on the subject's previous action against A, independent of action against B. In each trial (Fig. 1*A*, test), one of the two computer agents (A or B) was presented in pseudorandom order and symbolized by a different neutral human face (Ogawa and Oda, 1998). Different pairs of faces were used for every subject to avoid directing brain activity to specific faces. The subjects were informed that their opponents were computer agents, that the time-invariant strategies of each agent were independent of each other, and they might be stochastic or might depend

Received June 11, 2008; revised March 9, 2009; accepted March 11, 2009.

This research was supported by the Strategic Research Program for Brain Sciences. We thank K. Toyama, C. Frith, K. Sakai, T. Yamagishi, and K. Nakamura for helpful discussions and S. Tada for technical support.

Correspondence should be addressed to Dr. Masahiko Haruno, ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Soraku-gun, Kyoto 619-0288, Japan. E-mail: mharuno@atr.jp.

DOI:10.1523/JNEUROSCI.2707-08.2009



Agent / Subject	Defect	Cooperate
Defect	-20	+50
Cooperate	-30	+30

Block	1						2						3						...	...	12					
	TEST			CONTROL			TEST			CONTROL			TEST			CONTROL			...	...	TEST			CONTROL		
Trial	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	...	...	1	2	3	1	2	3

Figure 1. Design of prisoner's dilemma task. ***A***, Timeline of test trials. After each test block, a control block was interleaved. Time schedule of control condition was exactly identical as test condition. ***B***, Payoff matrix. ***C***, Schedule of entire task.

on the subject's behavior. Therefore, they fully understood that the computer agents were nonhuman without intention or emotion.

Subjects were instructed to decide cooperate or defect immediately after the agent's face appeared on the screen and to press the corresponding button at a "beep" presented 6 ± 0.5 s after the agent's face. Subject and agent actions were shown on the screen 7 ± 0.5 s later, as well as that trial's monetary reward. The actual monetary reward received per trial depended on the behavior of both the subject and the agent (Fig. 1*B*). When the subject cooperated, he received 30 yen if the agent also cooperated, but lost 30 yen if the agent defected. On the other hand, when the subject defected, the reward was 50 yen if the agent cooperated but a penalty of 20 yen if the agent defected. Therefore, the optimal behaviors for the subjects were cooperating with agent A and defecting for agent B. The subjects were told that their objective was to maximize their total reward by learning the optimal behavior for dealing with each agent. At the experiment's end, the subject received the total monetary reward won during the prisoner's dilemma task. Each test block consisted of three trials. A control task was interleaved (Fig. 1*A*,*C*) in which the subject passively pushed the same three buttons as in the preceding test block guided by visual instructions with a yellow circle (corresponding test face was also shown at the beginning) without a reward. Twelve sets of test (18 trials for both agents A and B) block–control blocks were presented to each subject (Fig. 1*C*).

MRI acquisition and preprocessing. The subjects ($n = 32$) were postgraduate students of Kyoto University and the Nara Advanced Institute of Science and Technology. Informed consent was obtained, and the protocol was approved by our institution's ethics committee. MRI scanning was conducted with a 1.5 tesla Marconi scanner at ATR Brain Activity Imaging Center (ATR-BAIC) (TR 2.5 s, TE 49 ms, flip angle 80°, FOV 192 mm, resolution 3 × 3 × 5 mm). High-resolution (T1 [1 × 1 × 1 mm] and T2 [0.75 × 0.75 × 5 mm]) structure images were also acquired from each subject. Before statistical analysis, we performed motion correction and nonlinear transformation into the standard space of the MNI coordinates. These normalized images were resliced into 2 × 2 × 2 mm voxels using the T2 template of SPM2 (Friston et al., 1995) and then smoothed with an 8 mm full-width half-maximum isotropic Gaussian kernel.

Learning model. We used a simple reinforcement learning model (*Q*-learning) (Sutton and Barto, 1998) to examine each subject's learning process. In each trial, we assumed that the subjects would predict their reward while playing against the agent based on the subject's previous and subsequent actions. Therefore, a subject's reward prediction (RP) in the *t*th trial was denoted as $Q_t(\text{ag}, a^s_{t-1}, a^s_t)$, the predicted reward based on agent ag (agents A or B), previous behavior a^s_{t-1}, and subsequent behavior a^s_t (cooperate or defect) of the subject. We included a^s_{t-1} because this information is essential for predicting and learning against agent A ("tit-for-tat"). We expected the subjects to choose action a^s_t (cooperate or defect) with a larger RP and to learn by updating RP in proportion to the reward prediction error, i.e., the difference between RP and the actual reward (r_t). Therefore, the subject's learning process (model I) is simulated by the following equation using subject's actual action a^s_t, which reduces the reward prediction error for the next occurrence of the same combination of agent and actions: if the model correctly selects the subject's action by taking a larger RP (Q),

$$Q_t(\text{ag},a^s_{t-1},a^s_t) = Q_t(\text{ag},a^s_{t-1},a^s_t) + \alpha_t(r_t - Q_t(\text{ag},a^s_{t-1},a^s_t)),$$

else

$$Q_t(\text{ag},a^s_{t-1},a^s_t) = Q_t(\text{ag},a^s_{t-1},a^s_t) + \alpha_t(r_t - Q_t(\text{ag},a^s_{t-1},a^s_t)),$$

$$Q_t(\text{ag},a^s_{t-1},\bar{a}^s_t) = Q_t(\text{ag},a^s_{t-1},\bar{a}^s_t)\text{-}\Delta$$

end.

If the model cannot correctly select the subject's action, the second equation is applied to avoid situations where Q for incorrect action ($\bar{a}^s_t$) is larger than its true value. Δ was set to 5.0 in all simulations because the Δ value between 2.5 and 10.0 made no significant difference ($p < 0.05$; *t* test) in the prediction power of the subject's behavior. However, unstable learning was seen when the value was selected outside this range. Considering the agent's previous action, $Q_t(\text{ag}, a^{\text{ag}}_{t-1}, a^s_t)$ also produced results comparable to $Q_t(\text{ag}, a^s_{t-1}, a^s_t)$ in subsequent analysis. We also tested an even simpler model (model II) that does not consider the subject's previous action (a^s_{t-1}) [i.e., $Q_t(\text{ag}, a^s_t)$]. All elements of $Q_t(\text{ag}, a^s_{t-1}, a^s_t)$ were initially set at 0. Learning rate α_t, which controls the amplitude of the change, was determined by a standard recursive least-square procedure with an initial value of 100 (Haruno and Kawato, 2006). This simple learning procedure was expected to capture both mentalizing-based and simple reinforcement learning since subjects are grouped based on their different behaviors, as detailed later. The *Q*-learning model approximates each subject's trial-based learning curve, and we do not claim that it is in the brain in exactly the same form.

Statistical analysis of fMRI data. The fMRI data were analyzed using standard procedures for random effect models (i.e., one- or two-sample *t* tests) in SPM2 (Friston et al., 1995). We included three simple events in our regressors: agent presentation, beep sound, and reward feedback. We used each subject's RP as a parametric modulator at the agent presentation event. Hereafter, we focus on the RP results because we are mainly interested in interactive reinforcement learning. The imaging results did not change even if we included RP error as another regressor at the timing of the reward feedback. The illustrations of the statistical maps (see Fig. 3) were prepared using "multi_color" (http://www.cns.atr.jp/multi_color/), our in-house software. In Figures 3*B* and 4*A*, each subject's BOLD signal was extracted by MarsBar (http://marsbar.sourceforge.net/), and the signal increase ratio was computed by subtracting the average over the whole sequence. BOLD data were high-pass filtered (cut-off frequency: 128 s) by the Butterworth filter in MATLAB.

Results

Behavioral results

Agent A (tit-for-tat) exhibits simple and typical interactive behavior, but agent B does not react to the subject's behavior.

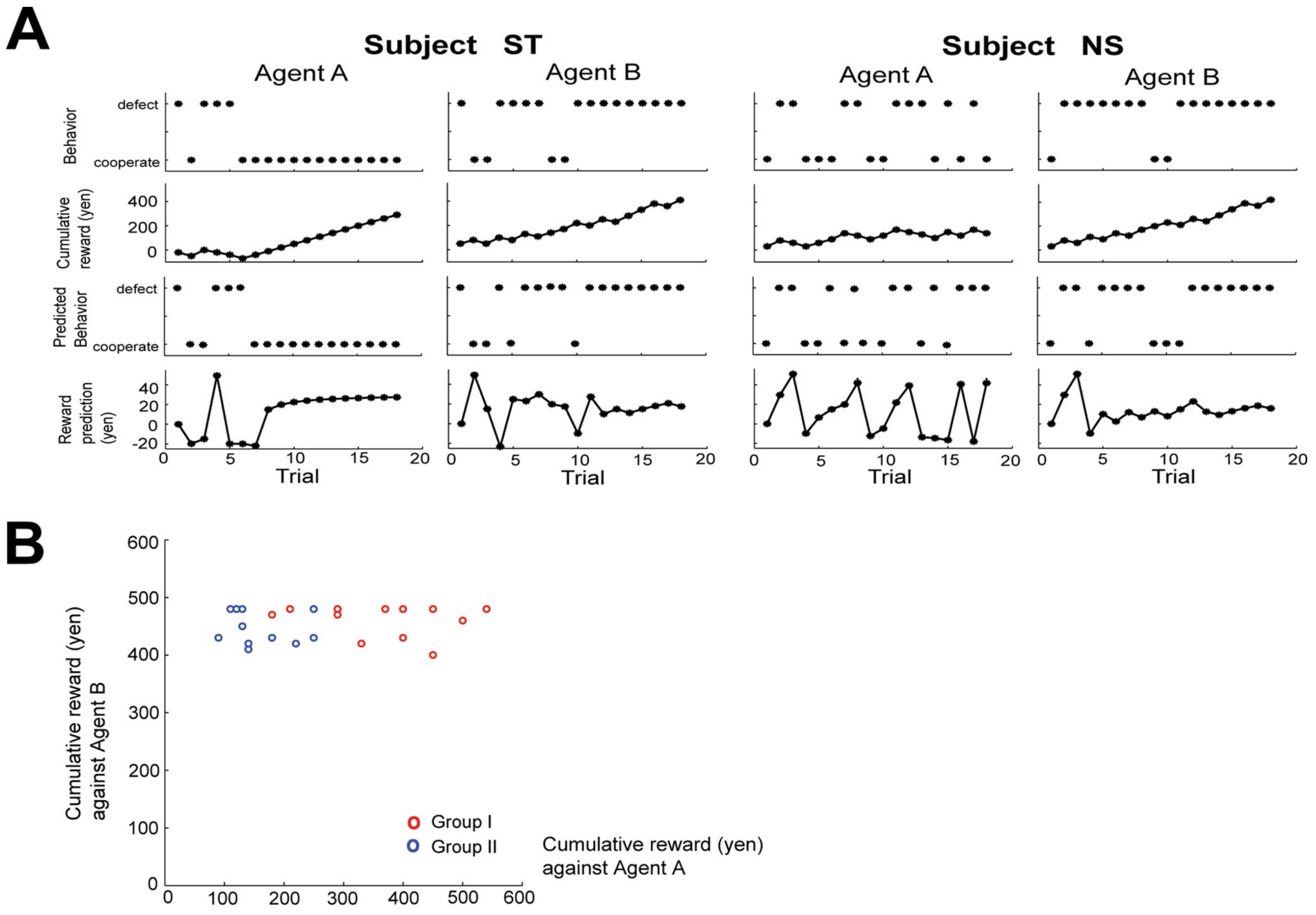


Figure 2. Behavioral data. ***A***, Behavior of two typical subjects. First two rows from top illustrate actions taken by subject and cumulative reward. Bottom two rows plot actions predicted by the model and estimated reward prediction during series of trials. ***B***, Cumulative reward of subjects. Cumulative reward obtained against agents A and B is plotted for group I in red and in blue for group II.

Therefore, to learn to interact with agent A, the subject is expected to consider the other agent's action strategy more when selecting an action. On the other hand, simple association of the subject's own action and reward was sufficient to learn against agent B. We therefore expected that more subjects would learn the optimal strategy for dealing with agent B than A (i.e., cooperating with agent A and defecting for agent B). The evaluation of subject performances was based on whether they selected the optimal behaviors in all of the last four trials for both agents. Among the 32 subjects, 12 learned the optimal behaviors against both agents A and B and were placed in group I. Twelve other subjects only learned the optimal behavior for dealing with agent B and were placed in group II. This categorization of subjects (and therefore subsequent imaging results) was unchanged even when the evaluation criterion was based on the last five trials. Gender and age were comparable between groups I (eight males, four females, mean age ± SD: 25.2 ± 2.04 years) and II (eight males, four females, mean age ± SD: 24.8 ± 1.55 years). The remaining eight subjects (five males, three females, mean age ± SD: 26.1 ± 2.42 years; group III) did not learn the strategies of agents A and B. After completing the prisoner's dilemma task in the scanner, the experimenter interviewed them and asked whether they understood each agent's strategy. All group I subjects correctly identified the strategies of both agents A and B, but in group II, only four and eight subjects identified the strategies of agents A and B, respectively. We judged that the subjects successfully understood each agent's strategy when "repeating the subject's previous action" and "random cooperation with a considerable probability" were implied in their statements for agents A and B.

In Figure 2*A*, the first and second rows show the chronological plots of the behavior selections and the cumulative rewards for each agent during learning by representative subjects in groups I and II. Subject ST in group I initially fluctuated between defect and cooperate when dealing with agent A before discovering around the sixth trial that cooperating was the optimal solution. The same subject quickly learned that the optimal strategy against agent B was to defect, although he/she tried another option at the eighth and ninth trials. After learning the optimal strategy, the subject's cumulative reward steadily increased for both agents. In contrast, subject NS in group II did not show convergence toward any fixed strategy against agent A and displayed wider behavior fluctuations than subject ST, but learned the optimal behavior against agent B after a few trials. Figure 2*B* plots the total amount of reward obtained by each subject in groups I (red) and II (blue) while playing agents A and B. The amount of the minimum (maximum) and mean (SD) total reward for groups I and II were [agent A: 180 (540) and 367.5 (111.3), agent B: 400 (480) and 460.8 (28.1) yen] and [agent A: 90 (250) and 156.6 (55.0), agent B: 410 (480) and 449.2 (28.7) yen], respectively. In accordance with the group definitions, the reward against agent A was significantly larger in group I than in II (p = 0.000064; t test), and the difference between the groups when playing against agent B was not significant (p = 0.33; t test).

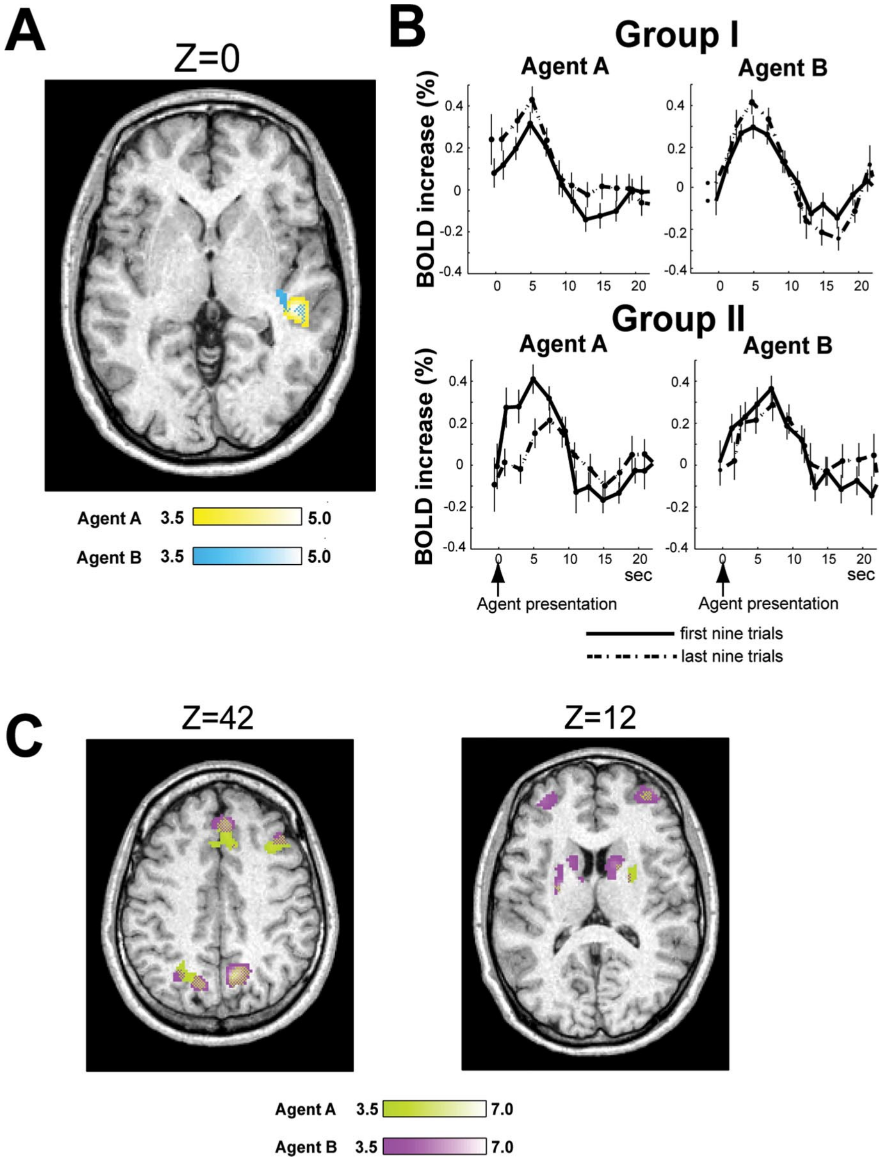


Figure 3. Correlation between fMRI data and RP. ***A***, Differential correlation between groups I and II. Significant difference between groups I and II only found in STS, yellow against agent A and light blue against agent B, with overlap depicted as a mosaic. Brightness of color bars reflects *T* values. MNI coordinates of peak-correlated voxels for agents A and B were [48, −30, 0] and [48, −32, 2], respectively. ***B***, Learning-related BOLD signal change of group I and group II subjects in STS peak voxels. Error bars represent SE. Statistical test was done at 5 s after presentation of agent ($p < 0.05$; *t* test uncorrected for multiple comparisons in selecting the voxel). ***C***, Brain activity commonly correlated with RP in groups I and II. Horizontal sections of brain areas are shown. MNI Z coordinates of slices depicted were 42 and 12. *T* values are light green for agent A and magenta for agent B.

Furthermore, the mean (SD) probability of choosing defects over all 18 trials against agent A were 0.22 (0.14) in group I and 0.54 (0.11) for group II. There was no significant difference between groups I and II in mean reaction time after the triggering beep against either agent A ($p = 0.98$; *t* test) or B ($p = 0.66$; *t* test).

To further examine each subject's learning process, we used a simple reinforcement learning model that considered the subject's previous and subsequent actions (model I). In Figure 2*A*, the third and fourth rows show action predicted by the model and reward prediction (RP) estimates for subjects ST and NS, respectively. Against agent A, the RP of subject ST fluctuated during trial and error and converged around the eighth trial. Subject ST also showed rapid RP convergence in relation to agent B, although fluctuations due to the stochastic nature of agent B were also visible. In contrast, subject NS showed larger RP fluctuations against agent A than subject ST, but showed a similar RP pattern against agent B to subject ST.

We evaluated this model based on how precisely it reproduced the subject's behavior: cooperate or defect. The mean accuracy over subjects and the SD of the current model (model I) were the following: group I ($n = 12$): 0.87 ± 0.11 against agent A, 0.93 ± 0.10 against agent B; group II ($n = 12$): 0.67 ± 0.19 against agent A, 0.87 ± 0.13 against agent B; other subjects ($n = 8$): 0.60 ± 0.097 against agent A, 0.76 ± 0.16 against agent B. We also examined a simpler model (model II) that only considered the subject's next action (group I: 0.77 ± 0.13 against agent A, 0.94 ± 0.073 against agent B; group II: 0.61 ± 0.11 against agent A, 0.84 ± 0.19 against agent B; other subjects: 0.60 ± 0.10 against agent A, 0.76 ± 0.16 against agent B). Group I subject behavior only against agent A was significantly better reproduced by model I than model II in the percentage of choice explained ($p < 0.05$; *t* test). These results show that without taking account of subject's previous behavior, the *Q*-learning model defected more against agent A (tit-for-tat) for group I subjects. Such a difference was not seen against agent B, and no significant difference was found between models I and II for group II subjects. All these observations indicate a key role of a^s_{t-1} for group I in dealing with agent A (not the general over-fitting due to this additional parameter), and are consistent with the view that group I subjects take a model-based strategy to predict the opponent's behavior based on their previous behaviors.

Imaging results

To identify the differences in learning-related neural activity between groups I and II, we performed linear correlation analysis of the fMRI data with RP by the contrast on the mean betas for the RP regressor and contrasted the two groups. Figure 3*A* shows the differential neural correlates between groups I and II ($p < 0.001$, uncorrected, and cluster size >10). The superior temporal sulcus (STS) was the only brain structure where group I showed a statistically greater significant correlation against agent A (yellow) than group II. Importantly, the same analysis against agent B (light blue) also only lit up the STS with the same statistical threshold, and there was substantial overlap with the results for agent A. This consistency in the results between agents A and B negates the possibility that STS activity arose from a difference between the two groups in behavior or model fitness against agent A because they behaved similarly against agent B. The opposite contrast detected no significant differences for either agents A or B. We also conducted similar correlation analysis using model II without considering previous subject behavior and found no such difference in the STS activity against agent A, even with a very low threshold ($p <$

0.01); the result against agent B remained similar. Additionally, to eliminate the possibility that this STS activity arose from sensorimotor processes and implicit emotions for face perception (Singer et al., 2004), we conducted the same correlation analysis of the control trials by using RP of corresponding test trials and found no difference in STS activity between groups I and II, even with a very low threshold ($p < 0.1$; uncorrected).

Figure 3*B* displays how the BOLD signal changes during learning in the STS peak voxel of the group I and II subjects. It was separately averaged over the subjects for agents A and B, and for the first nine trials (former half of all 18 trials, bold lines) and the last nine trials (latter half, dotted lines). STS activity in the group I subjects increased in the later trials for both agents A and B ($p <$ 0.05 at 5 s after agent presentation; *t* test uncorrected for multiple comparisons in selecting the voxel), while group II subjects did not show this tendency. Thus, STS activity increased during learning in the group I subjects, but not in the group II subjects. Because RP is expected to represent the reward prediction based on the other agent's strategy, the RP of the group I subjects takes higher values and is used more often in the late phase of learning than the beginning. The STS activity of the group I subjects was larger in the second half of the trials than the first, consistent with the behavior of RP.

Figure 3*C* shows the brain areas commonly activated in both groups I and II in correlation with RP with identical statistical thresholds in Figure 3*A*. The activities in the dorsolateral and ventral prefrontal, the anterior cingulate, the parietal cortices, and the striatum showed a statistically significant correlation with the complete overlap between agents A (green) and B (magenta). The same RP analysis conducted for the eight subjects (group III) not in groups I or II lit up almost the same areas as those activated in groups I and II (Fig. 3*C*), although rigorous comparison of brain activity is susceptible to behavioral differences. Specifically, compared with group II subjects, the only noticeable difference was that group II showed larger correlation of activity in the caudate nucleus for both agents A and B than group III with a moderate threshold ($p < 0.005$, uncorrected, and cluster size >10). Importantly, the STS activity did not exhibit differences between groups II and III for either agents A or B, even with a very low threshold ($p < 0.1$).

The STS activity in Figure 3*A* is a within-subject finding because each group I subject showed increased activity in the STS correlated with that subject's evaluation of learning performance (RP). An intriguing question here is whether STS activity can evaluate a subject's learning performance in comparison with other subjects. Figure 4*A* plots the relationship between each subject's average reward for each agent in the late phase of learning and signal increase in the STS (Fig. 3*A*, peak voxels) in those trials upon presentation of the agent (group I: top; group II: lower). In group I there was a statistically significant positive correlation in relation to agents A ($p = 0.024$) and B ($p = 0.025$), but not in group II. It would be noteworthy that in the peak selection (Fig. 3), we used the contrast on the mean beta values for the RP regressor, while in the subsequent across-subject analysis (Fig. 4*A*), we conducted correlation analysis of each subject's event-triggered activity in the peak voxels with the subject's behavioral performance (i.e., average reward over last 10 trials). Thus, since the two correlation measures are independent, the result in Figure 4*A* is likely to provide unbiased correlation. Figure 4*B* plots the point at which each group I subject learned the optimal behavior and the subject's average reward, demonstrating that subjects who learned earlier obtained more reward against both agents A ($p = 0.002$) and B ($p = 0.014$) and engaged their STS more often.

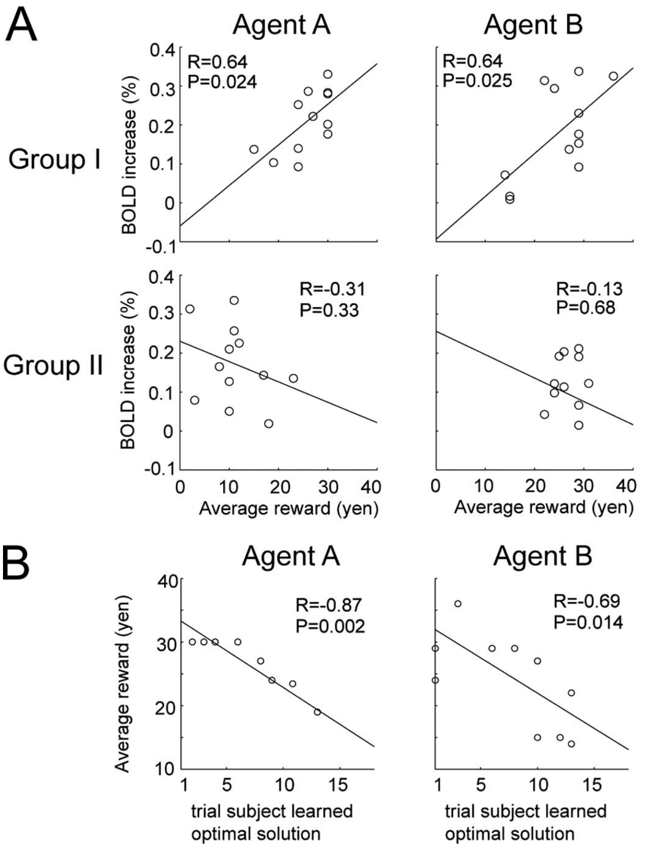


Figure 4. Relationship between each subject's learning performance and STS activity. ***A***, Each panel plots average reward over last 10 trials for each agent and corresponding increase in BOLD signal in STS averaged for 7.5 s after presentation of each agent. BOLD signal was taken from peak voxels revealed by previous analysis [i.e., [48, −30, 0] and [48, −32, 2], respectively (Fig. 3*A*)]. ***B***, Relationship between the point when each group I subject learned the optimal solution against each agent and the subject's average reward over last 10 trials.

Discussion

The current study showed that group I subjects exhibited an increase in STS activity during interactive reinforcement learning, but not in the group II subjects, although reward-related brain structures were similarly recruited by both groups. This STS activity was also predictive of learning performances across the group I subjects. Attributing this differential activity to a behavioral difference is difficult since it persisted in relation to agent B.

Together with the orbitofrontal and paracingulate cortices, STS has been implicated in "the theory of mind" or "mentalizing" (Baron-Cohen, 1997; Frith and Frith, 2003) during human-human interactions. Therefore, the involvement of STS during interaction games is a reasonable consequence (Rilling et al., 2004; Hampton et al., 2008). Hampton et al. recently reported that STS activity was correlated with an update (error) signal at the timing of the reward feedback. Compared with these studies, we showed that the learning signal in STS at the timing of the agent presentation was predictive of individual differences in the learning competence of group I subjects, but not of group II subjects, which is consistent with mentalizing (Singer et al., 2004). We also showed that this STS activity was found through interaction with nonhuman agents. A future experiment that incorporates explicit modeling of mentalizing and reward processing might help extend our results.

We speculate that the orbitofrontal and paracingulate cortices did not show up (Fig. 3*A*) because STS mainly provides predictive knowledge of the other agent's behavior (Frith, 2007) for reward-action association, while the other two areas are involved in the estimation and evaluation of the other human agent's intention, which were unnecessary in our task. Correspondingly, STS activity in Figure 3*A* was located slightly anterior to the typical "theory of mind" area (Singer et al., 2004) and toward the locus of biological motion perception and prediction (e.g., mouth) (Pelphery et al., 2005).

A third of the group II subjects accurately reported agent A's tit-for-tat strategy in postexperiment interviews but could not behave optimally, suggesting that STS also plays a role in linking the knowledge of the other agent's behavior and reward-based action selection. The finding that the two groups similarly recruited brain structures for reward processing such as the striatum (Schultz et al., 2003; Haruno et al., 2004; O'Doherty et al., 2004; Haruno and Kawato, 2006), prefrontal (Barraclough et al., 2004), anterior cingulate (Kennerley et al., 2006; Seo and Lee, 2007), and posterior cortices (Dorris and Glimcher, 2004) is consistent with this view (Fig. 3*C*).

Our behavioral and modeling results indicate that simple association between subject's subsequent action and reward was not sufficient to explain group I subject behaviors against agent A. However, simple association did work well to explain group I subject behaviors against agent B and group II subject behaviors against both agents. In contrast, the imaging results showed that group I subjects recruited STS against both agents A and B, while group II subjects did neither. This difference suggests that group I subjects were considering the other agent strategies, even though unnecessary for optimal behavior, while group II subjects were not or their efforts did not improve learning performance. Group I subjects apparently always used this STS function to boost the association between their actions and rewards, while group II subjects tended to rely on simpler associations. The difference in learning strategies between groups I and II only surfaced when playing agent A.

References

Axelrod RM (1984) The evolution of cooperation. New York: Basic Books.

Baron-Cohen S (1997) Mindblindness: an essay on autism and theory of mind (learning, development and conceptual change). Cambridge, MA: MIT.

Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7:404–410.

Camerer CF (2003) Behavioral game theory. New Jersey: Princeton UP.

Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron 44:365–378.

Fogg BJ (2003) Persuasive technology. San Francisco: Morgan Kaufmann.

Friston KJ, Holmes AP, Worsley K, Poline JB, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional brain imaging: a general linear approach. Hum Brain Mapp 2:189–210.

Frith CD (2007) Making up the mind. Oxford: Blackwell.

Frith U, Frith CD (2003) Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci 358:459–473.

Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A 105:6741–6746.

Haruno M, Kawato M (2005) Two groups of subjects with different learning competence in a prisoner's dilemma task exhibit differential activations in the superior temporal sulcus. Soc Neurosci Abstr 35:409.23.

Haruno M, Kawato M (2006) Different neural correlates of reward expectation and reward expectation error in putamen and caudate nucleus during stimulus-action-reward association learning. J Neurophysiol 95:948–959.

Haruno M, Wolpert DM, Kawato M (2001) Mosaic model for sensorimotor learning and control. Neural Comput 13:2201–2220.

Haruno M, Kuroda T, Doya K, Toyama K, Kimura M, Samejima K, Imamizu H, Kawato M (2004) A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. J Neurosci 24:1660–1665.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947.

Lee D (2008) Game theory and neural basis of social decision making. Nat Neurosci 11:404–409.

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

Ogawa N, Oda M (1998) Design and evaluation of facial expression database. ATR Technical Report TR-H-244.

Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G (2005) Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. Cereb Cortex 15:1866–1876.

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) The neural correlates of theory of mind within interpersonal interactions. Neuroimage 22:1694–1703.

Schultz W, Tremblay L, Hollerman JR (2003) Changes in behavior-related neuronal activity in the striatum during learning. Trends Neurosci 26:321–328.

Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci 27:8366–8377.

Singer T, Kiebel SJ, Winston JS, Dolan RJ, Frith CD (2004) Brain responses to the acquired moral status of faces. Neuron 41:653–662.

Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT.

Van Lange PAM (1999) The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. J Pers Soc Psychol 77:337–349.