Cortex

CrossMark

Special section: Research report

Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain

Hideyuki Takahashi [a,b,c], *Kazunori Terada* [d], *Tomoyo Morita* [a,c], *Shinsuke Suzuki* [e,f], *Tomoki Haji* [b,c], *Hideki Kozima* [g], *Masahiro Yoshikawa* [h], *Yoshio Matsumoto* [i], *Takashi Omori* [b], *Minoru Asada* [a] and *Eiichi Naito* [c,j,*]

[a] *Graduate School of Engineering, Osaka University, Japan*
[b] *Brain Science Institute, Tamagawa University, Japan*
[c] *CiNet, NICT, Japan*
[d] *Department of Information Science, Gifu University, Japan*
[e] *Division of the Humanities and Social Sciences, California Institute of Technology, United States*
[f] *Graduate School of Letters, Hokkaido University, Japan*
[g] *School of Project Design, Miyagi University, Japan*
[h] *NAIST, Japan*
[i] *AIST, Japan*
[j] *Graduate School of Medicine, Osaka University, Japan*

ARTICLE INFO

ABSTRACT

Internal (neuronal) representations in the brain are modified by our experiences, and this phenomenon is not unique to sensory and motor systems. Here, we show that different impressions obtained through social interaction with a variety of agents uniquely modulate activity of dorsal and ventral pathways of the brain network that mediates human social behavior.

We scanned brain activity with functional magnetic resonance imaging (fMRI) in 16 healthy volunteers when they performed a simple matching-pennies game with a human, human-like android, mechanical robot, interactive robot, and a computer. Before playing this game in the scanner, participants experienced social interactions with each opponent separately and scored their initial impressions using two questionnaires.

We found that the participants perceived opponents in two mental dimensions: one represented "mind-holderness" in which participants attributed anthropomorphic impressions to some of the opponents that had mental functions, while the other dimension represented "mind-readerness" in which participants characterized opponents as intelligent. Interestingly, this "mind-readerness" dimension correlated to participants frequently

---

* *Corresponding author.* Center for Information and Neural Networks (CiNet), National Institute of Information and Communication Technology (NICT), 2A6, 1-4 Yamadaoka, Suita, Osaka 565-0871, Japan.
  E-mail address: eiichi.naito@nict.go.jp (E. Naito).

changing their game tactic to prevent opponents from envisioning their strategy, and this was corroborated by increased entropy during the game. We also found that the two factors separately modulated activity in distinct social brain regions. Specifically, mind-holderness modulated activity in the dorsal aspect of the temporoparietal junction (TPJ) and medial prefrontal and posterior paracingulate cortices, while mind-readerness modulated activity in the ventral aspect of TPJ and the temporal pole.

These results clearly demonstrate that activity in social brain networks is modulated through pre-scanning experiences of social interaction with a variety of agents. Furthermore, our findings elucidated the existence of two distinct functional networks in the social human brain. Social interaction with anthropomorphic or intelligent-looking agents may distinctly shape the internal representation of our social brain, which may in turn determine how we behave for various agents that we encounter in our society.

1. Introduction

It is well established that internal (neuronal) representations in the brain can be modified by experience. Many previous studies have elucidated how our sensory and motor experiences shape representations in sensory-motor systems; however, this phenomenon is not limited to the fundamental systems since modifications can also be observed in brain networks that mediate social behavior.

For example, when we play a game with others, we often change our tactics depending on strategies used by our opponents. In this way, our brain flexibly modifies our attitudes and actions based on the perception and interpretation of our opponent's behavior (Delgado, Frank, & Phelps, 2005; Frank, Gilovich, & Regan, 1993; Kuzmanovic et al., 2012; Parise, Kiesler, Sproull, & Waters, 1999).

The brain network that mediates social interaction consists of the posterior end of the superior temporal sulcus (pSTS), the adjacent temporoparietal junction (TPJ), the temporal pole (TP), the medial prefrontal cortex (mPFC), and the posterior paracingulate cortex (PCC) (Frith & Frith, 1999, 2003, 2006; Olson, Plotzker, & Ezzyat, 2007; Saxe, 2006). Many previous studies suggest that these brain regions are assigned specific roles. For example, the mPFC and PCC are mainly activated when an individual meditates his/her own mental state or when they infer another individual's mental state (Amodio & Frith, 2006; Frith & Frith, 1999, 2003, 2006). On the other hand, the TP seems to be involved in more emotional aspects of social processing. Specifically, it has been proposed that this region relates social perception with emotion since dysfunction of this region leads to various psychiatric disorders related to emotional regulation (Olson et al., 2007). Finally, the TPJ/pSTS seems to cover a wide range of socio-cognitive functions, such as social perception, perspective taking, and theory of mind.

Modulation of the activity in these social brain networks is known to be dependent on various factors including interaction between other humans and even robots (Chaminade et al., 2012; Krach et al., 2008). Thus, it is likely that activity in our social brain network is uniquely modulated by how we perceive and interpret the structurally complex characteristics of others (Gray, Gray, & Wegner, 2007; Haslam, 2006; Loughnan & Haslam, 2007); however, this has yet to be fully elucidated. To address this question, we prepared different types of agents including human-like and non-human-like robots since multiple agents likely give us different impressions due to their specific characteristics. Indeed, neuroimaging research on our interactions with non-human agents (android, robot, and artificial intelligence; Chaminade et al., 2012; Krach et al., 2008) may help us to understand how our brain forms an internal representation of social interactions.

In the present study, we designed a simple matching-pennies game and performed functional magnetic resonance imaging (fMRI) on 16 healthy volunteers while they played against five different types of opponents: human, human-like android, mechanical robot, interactive robot, and a computer. As described, we included the robots in order to manipulate the degree of human-like appearance and attitude, which could elicit unique impressions to the participants who were naïve to robots. Before playing this game in the fMRI scanner, participants had a chance to socially interact with each opponent and scored their impressions about each interaction using a questionnaire. We expected that participants would have multi-dimensional perceptions about each opponent's characteristics since this has been indicated by previous studies (Gray et al., 2007; Haslam, 2006; Loughnan & Haslam, 2007). We then tested our hypothesis that participants change their behaviors (tactics) in subsequent gameplay depending on their putative multi-dimensional perceptions about each opponent obtained through the pre-scanning social interaction. In addition, we hypothesized that the participants' multi-dimensional perceptions would have a unique impact on different sets of social brain networks during gameplay (Fukui et al., 2006; Gallagher, Jack, Roepstorff, & Frith, 2002; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004).

2. Methods

2.1. Participants

In total, 20 healthy, right-handed volunteers participated in this study. None of the participants had a history of neurological or psychiatric illness. All participants provided written

informed consent prior to the onset of this study, which was approved by the Ethics Committee of Tamagawa University, Japan. We analyzed the fMRI data obtained from 16 participants (five males; age range, 18—25 years), and excluded the data obtained from the remaining four participants due to excessive head motion (>5 mm) during the scan.

2.2. Tasks

2.2.1. General procedure

We scanned brain activity using fMRI while participants played a matching-pennies game with each of the five opponents. The opponents included a human (woman in her 20 sec), a human-like android (Actroid F; Yoshikawa, Matsumoto, Sumitani, & Ishiguro, 2011), a mechanical robot (Infanoid; Kozima, 2002), an interactive robot (Keepon; Kozima, Michalowski, & Nakagawa, 2008), and a computer (Fig. 1a). As for the human opponent, we enrolled four women in order to match the gender with female Actroid F, and in order to avoid a particular opponent giving a specific impression to the participants. Human opponents were randomized participant by participant, but a given participant always played the game against the same female opponent. All participants were completely naïve to Actroid F, Infanoid, Keepon, and the particular human opponent. Before participants played the game in the fMRI scanner, each participant experienced a short conversation with each opponent outside the scanner (Fig. 1b) and was asked to answer two questionnaires.

2.2.2. Short conversation outside the scanner

Each participant chatted with each opponent one-on-one for 30 sec prior to entering the scanner room. The order of the opponents was randomized across participants. Conversations included the following three defined topics: first, the participant verbally asked the name of the opponent, and then briefly described their impression of their opponent, followed by the participant talking about his/her ardor for the game against the opponent. All opponents reacted with verbal responses and/or bodily gestures that were consistent across participants (see Supplemental movies). When an opponent was human, Actroid F, or Infanoid, it reacted to the first question as follows: "My name is (name of the opponent). Nice to meet you," to the second question, "Thank you very much for your kindly comment," and to the third comment, "I will do my best too." Both the human and Actroid F opponents bowed

to the participant (in physical Japanese-style greeting) before and after the chatting session and did not move their bodies except for this action. The Infanoid moved its head in order to track the participant's face and moved its hands arbitrarily during chatting. The Keepon did not talk and simply reacted to the participant's speech by wiggling its body. In the case of the computer that could not make bodily gestures, the participant could only see the flow of complex program code in the monitor, which we expected would give an impression of intelligence to the participant.

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.cortex.2014.03.011.

Immediately after the interaction with each opponent, participants were asked to fill out questionnaires about each opponent. The "impression questionnaire" required the participant to rate the impressions of the opponent and was a modified version of an original Japanese questionnaire (Kanda, Ishiguro, Ono, Imai, & Nakatsu, 2002) that included 22 adjective items (human-like, intelligent, ethical, nice, cute, friendly, active, positive, kind, warm, curious, thoughtful, emotionally stable, rational, responsible, biological, conscious, regular, natural, simple, emotional). Participants were told to rate each opponent based on how well each adjective item described the each opponent by choosing a number from 1 to 7, with 1 indicating "an item does not fit to the character of the opponent at all" and 7 indicating "an item fits very well to the character of the opponent."

The "mental function questionnaire" consisted of nine sentences (#1—#9) as we listed in Supplemental material (e.g., "When I ignore the opponent, it will appeal to direct my attention toward it."). We asked participants to evaluate the likelihood of performing the behavior described in each sentence by choosing yes or no. When a participant thought that an opponent would likely perform the behavior described in a sentence, they chose yes, otherwise they selected no. Through the evaluation process using the second questionnaire, we could infer to what degree the participants explicitly attributed mental functions to each opponent, because not only a simple reaction of an opponent but also fundamental aspects of opponent's mentality, such as thinking (#1), inference (#9), emotion (#4 and #5) and motivation (#6), can be evaluated by some of the present sentences. Participants took approximately 5 min to complete the two questionnaires for each opponent.

Using this design, all participants experienced social interactions with all opponents and evaluated their impressions
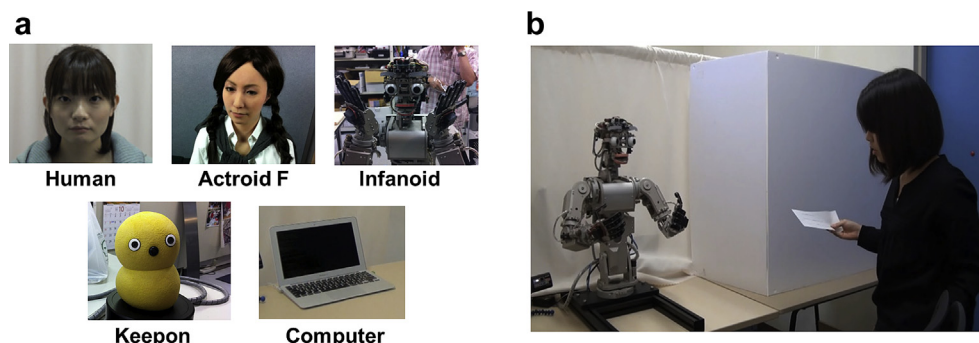
Human   **Actroid F**   **Infanoid**

Keepon   **Computer**

Fig. 1 — Five opponents (a) and a scene depicting a short-period conversation outside the scanner (b).

before beginning the game. Importantly, the impressions were not preconceived notions that the participants had formed before they met the opponents, since all of them were naïve to each agent, and the impressions emerged purely from perceptual dimension of the participants, which were not corroborated by any particular substantial mental faculties of the opponents.

2.2.3.   Matching-pennies game

After participants completed the questionnaires for all opponents, they were asked to lie in the fMRI scanner with their heads immobilized with an elastic band and sponge cushions and their ears plugged. Visual stimuli were presented on a projection screen and viewed by the participants via a mirror mounted on the head coil. Inside the scanner, the participant played a matching-pennies game which is often cited in game theory literature as the most simplified example of a zero-sum competitive situation. The game was played between two players: a participant and an opponent. In each trial, players were requested to select either the left or the right side of their bodies, with each participant making his/her decision by pressing a button. A win or loss was determined based on the combination of their decisions. Only when each participant selected the same side as the opponent, he/she was awarded with 100 yen (about 1 US$), otherwise he/she lost 100 yen. Thus, being able to predict an opponent's thoughts was very important to win the game. The two leftward and rightward arrows in the "Outcome" panel of Fig. 2 indicate selected directions by the opponent (leftward) and by the participant (rightward), respectively. In the example presented in Fig. 2, the participant lost 100 yen.

Before participants entered the scanner room, they were given instructions on how to play the game and were encouraged to accumulate as many wins as possible. Importantly, participants were also instructed that they would play the game against each opponent, which was presented as a small icon at the top of the monitor inside the scanner. We informed participants about the possibility that different opponents may use different game tactics by providing them with instructions stating that, "considering characteristic differences among opponents may increase the reward that you receive." By giving this instruction, we expected that participant's attitude specific to each opponent could become prominent. All participants completed five practice trials with a computer opponent before entering the scanner room.

Once in the scanner, participants played against the same pre-programmed computer algorithm for all five opponents (human, Actroid F, Infanoid, Keepon, computer), which generated each of two directions (left or right) with an equal probability. Therefore, an expected wining ratio was fixed to 50% in each trial, regardless of the opponent. As confirmed by interviewing participants after the experiment, the opponent presented in the icon of each trial was believed to be the opponent during that session.

Each trial lasted for 2 sec (Fig. 2), and participants were required to select either the left or right side by pressing one of two buttons with their index or middle finger within 1 sec. The directions selected by the participant and by the opponent were indicated by arrows, which were displayed in the monitor so that the participant would know if he/she won or lost in each trial. This was presented for 1 sec, and when participants failed to respond within 1 sec, their response was randomly determined for that trial.

Each block consisted of 20 trials that lasted for 1 min. In each block, participants consistently played the game with the same opponent. We prepared 15 blocks in a 15-min run. Each participant played with the same, randomly selected opponent three times in each run. A break was taken between runs, and the participants completed two runs in the scanner. Each block began with the presentation of a 2-sec cue that directly indicated the opponent with a word. Immediately after the cue disappeared, participants started the game, which lasted for 40 sec per block, i.e., 20 trials × 2 sec. The total amount of money earned in each block was displayed for 2 sec after the
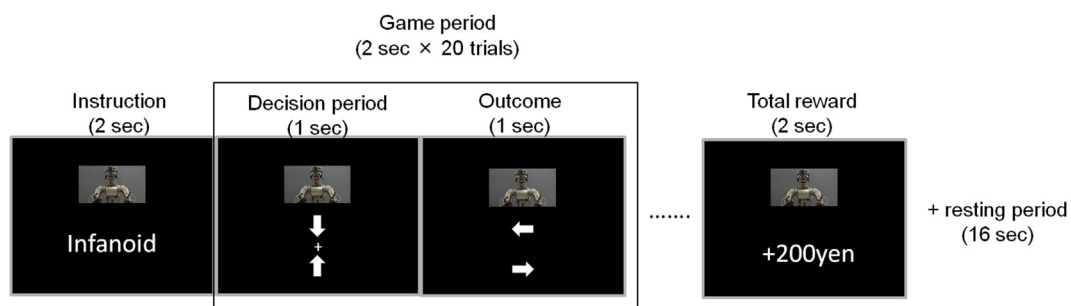
Fig. 2 — Time course of gameplay in a block. The series of panels represent a case of an Infanoid block, where a participant played the matching-pennies game with an Infanoid. Throughout a block, the opponent icon was consistently presented at the top of the screen. At the very beginning of a block, an opponent's name was presented (2 sec). Immediately after the game started, participants were required to select either left or right by pressing one of two buttons with their index or middle finger within 1 sec. During this decision period, upward and downward arrows were presented. Outcomes of decisions from the opponent and participant were shown as top and bottom arrows, respectively. The panel shows the case where the opponent chose its "right" and the participant chose his/her "right", meaning that the participant lost this game. This was repeated 20 times in each block. At the end of a block, the total reward was shown for 2 sec. In this case, the participant received a reward of 200 yen. There was a 16-sec resting period before the start of a next block where participants played with a new opponent.

game. We also included a 16-sec resting period at the very end of the block.

2.3. Data analysis

2.3.1. Behavioral analysis

We expected that participants would change their tactics depending on the various impressions they had formed of the different opponents they encountered during the game. One stringent way to evaluate the changes in tactics was to compute "entropy" (Ohira, Matsunaga, & Murakami, 2013; Takahashi, Saito, Okada, & Omori, 2013), which can be a measure of the degree of randomness or uncertainty in decision-making. Moreover, it has been previously shown that maximizing entropy is considered an optimal tactic in the matching-pennies game, according to game theory (Camerer, 2003; Nash, 1950). Therefore, a greater entropy value reflects difficulty in predicting a forthcoming participant's response based on the patterns of responses in previous trials. For example, if players followed a simple rule, such as a win-stay/lose-switch rule, the level of entropy tended to be low. In contrast, entropy was at a maximum when participants' behaviors were completely random during a trial.

We then quantified the randomness of decision-making for each block of 20 trials as entropy, H, which was calculated using the conditional frequency, $p(d|c)$, of the decision, d (L or R), selected in the current game context, c (the recent choices for participants and opponents). Entropy, H, indicated how the decision, d, was generated independently of the current game context, and the value of H positively correlated with the degree of randomness of decision-making in the matching-pennies game (Takahashi et al., 2013).

$p(d|c)$ was calculated from the following equation:

$$p(\mathrm{d}|\mathrm{s}) = \frac{n(\mathrm{d}|\mathrm{s}) + k}{\sum_i \{n(\mathrm{i}|\mathrm{s}) + k\}}$$

where a variable $n(d|c)$ indicates the number of times a decision d was made in the context of c, and k is a correction coefficient that prevents small samples from deforming $p(d|c)$. Due to the limitation of working memory capacity in humans, it is unlikely that participants were able to access the entire context in the game. Thus, we assumed that their decisions were made based on a portion of the context and assumed six partial contexts (pc) for entropy estimation (i.e., S1, the latest decision by the participant; S2, the last two decisions by the participant; O1, the latest decision by the opponent; O2, the last two decisions by the opponent; S1 & O1, a combination of the latest decisions both by the participant and by the opponent; none, no game context). Below, c_{pc} indicates the game context corresponding to each pc and entropy, $H(d|c_{pc})$, in each block was calculated using the following equation:

$$H_{pc} = -\frac{1}{N_{pc}} \sum_{c_{pc}} \sum_{d} p(\mathrm{d}|c_{pc}) \log_2 p(\mathrm{d}|c_{pc})$$

Here, N_{pc} is the number of possible alternatives for a particular c_{pc} and this variable normalizes H_{pc} in a range from 0 to 1. For each block, the lowest value among the six entropy values was chosen as the decision-entropy value for that block. Importantly, this value could potentially increase toward a value of one as decisions became less predictable.

We next calculated mean entropy for each opponent and each participant separately, and then calculated the grand mean across participants (Fig. 3). We performed a one-way analysis of variance (ANOVA) for entropy and post-hoc *t*-tests to evaluate the statistical difference in entropy between opponents.

2.3.2. Questionnaire analysis

We performed principle component analysis (PCA) for the results obtained from the impression questionnaire by pooling the data obtained from all participants. The factor structure of 21 items (see above) was assessed by PCA (Lisetti, Brown, Alvarez, & Marpaung, 2004) using the MATLAB statistical toolbox (The MathWorks Inc., Natick, MA). We found three representative orthogonal axes (from the first to third components, see Table 1). Each opponent had a specific value for each PCA component, and these were further used as parametric covariates in our subsequent fMRI analysis.

In order to analyze data obtained from the mental function questionnaire, we simply counted the number of "yes" answers collected from participants for all nine sentences assigned to describe each opponent. This value was determined as a mental function score, which may represent how much the participants explicitly attributed mental functions to each opponent. Next, we calculated the correlation coefficient between the values of the PCA component obtained from the two questionnaires across the five opponents. This was done separately for each PCA component. Similarly, we also calculated correlation coefficients between the PCA and entropy values for each participant. We then transformed the acquired correlation coefficients to z-scores within each participant. One-sample *t*-tests were then performed to evaluate if the correlation coefficients obtained from all participants were significantly different from 0. By evaluating the correlations, we were able to determine which PCA component better reflected the mental function score (how much the participants explicitly attributed mental functions to
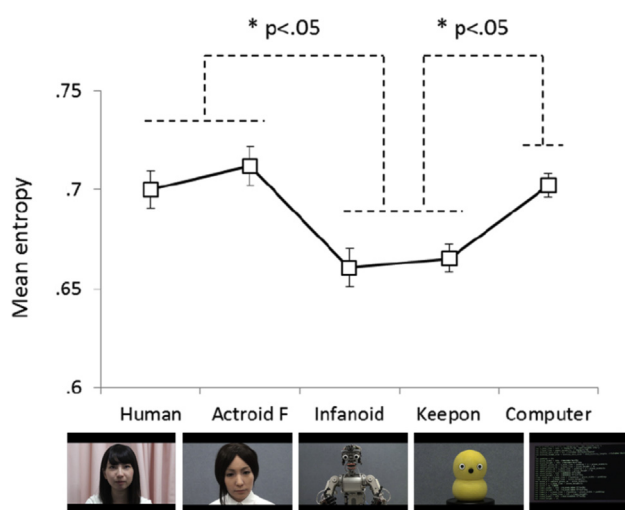
Fig. 3 — **Grand means of entropy for the five opponents across participants. Error bars indicate standard errors of means.**

Table 1 — Loads of questionaries' items for each PCA component.

	1st	2nd	3rd
Human-like	.3345	−.0081	.3498
Intelligent	.0745	.4607	.2512
Ethical	.0398	.4523	.1508
Nice	.1751	.0867	.0029
Cute	.2846	−.0619	−.281
Friendly	.3243	−.0007	−.3127
Active	.2078	.166	−.3477
Positive	.1671	.1705	−.3017
Kind	.19	.042	−.0562
Warm	.2748	.0066	−.1688
Curious	.1749	.0718	−.2756
Thoughtful	.164	.127	.0819
Emotionally stable	.0231	.348	−.0972
Rational	−.0748	.4174	−.0638
Responsible	.108	−.026	.2067
Biological	.3322	−.1024	.412
Conscious	.329	−.1184	.1665
Regular	−.1018	.0284	−.0017
Natural	.2841	.0633	.0812
Simple	.1265	−.4033	−.1303
Emotional	.2838	−.0642	.1071

opponents) and entropy (how unpredictably the participants changed their choices during the game).

2.4. fMRI scan

2.4.1. fMRI data acquisition
Functional imaging was conducted using a 3-T Siemens Trio A Tim MRI scanner. For functional imaging during the experimental sessions, interleaved T2*-weighted gradient-echo echo-planar imaging (EPI) sequences were used to acquire 44 continuous 3-mm-thick, trans-axial slices that covered nearly the entire cerebellum [repetition time (TR) = 3000 msec, echo time (TE) = 25 msec, flip angle (FA) = 90°, field of view (FOV) = 192 mm², 64 × 64 matrix, voxel dimensions = 3.0 × 3.0 × 3.0 mm]. A high-resolution anatomical T1-weighted image was also acquired for each participant. We collected 298 functional volumes in each 15-min run.

2.4.2. fMRI data pre-processing
In this analysis, we discarded the first four volumes to allow for magnetization equilibration. Data were then analyzed using Statistical Parametric Mapping 8 (SPM8, Wellcome Department of Cognitive Neurology, London, UK) software implemented in Matlab 2013a (The MathWorks, Inc.). After correcting for differences in slice timing within each image volume, head motion was corrected using the realignment program within SPM8. Following realignment, volumes were normalized to the Montreal Neurological Institute (MNI) space using a transformation matrix, which was obtained from the normalization process of the first EPI image of each participant to the EPI template. Finally, normalized fMRI data were spatially smoothed with an isotropic Gaussian kernel of 8 mm (full-width at half-maximum).

2.4.3. fMRI data analysis
We used a general linear model (GLM) to analyze the fMRI data. From the above-mentioned questionnaire data and

behavioral analyses, it was determined that participants most likely changed their game tactics based on the different impressions of their various opponents. Thus, we expected that these impressions formed before scanning would modulate brain activity during the game. We then prepared parametric regressors to depict such brain regions.

We prepared four regressors per participant: one regressor was game-related used to specify the game period composed of 20 consecutive trials in each block by excluding the last 16-sec resting period (see above), while the other three regressors, which were constructed based on the three PCA components obtained from the impression questionnaire analysis, were used for parametric modulation. Participants played with each opponent block by block and had specific values for each PCA component. Thus, we generated a parametric regressor by modulating the amplitude of the game-related regressor with the PCA value assigned to that opponent by the participant. This was done for all three PCA components.

The parametric modulation analysis for each PCA component was first performed in each participant separately. The result of this analysis was the estimated blood oxygen level-dependent (BOLD) signal change obtained from each of the 16 participants. To accommodate inter-participant variability, the images from all participants were entered into a random effects group analysis (second-level analysis; Friston, Holmes, & Worsley, 1999) using one-sample t-tests (15 degrees of freedom), and a voxel-wise threshold of $p < .001$ (uncorrected) was used to generate a cluster image. The significance of the cluster size was determined at $p < .05$ with the family-wise error rate (FWE) correction in the entire brain space.

3. Results

3.1. Impression of each opponent: results from PCA

The PCA analysis for the impression questionnaire revealed three representative orthogonal axes (from first to third components). The contribution rates of the first to third components were 74.4%, 13.2%, and 12.4%, respectively. The loads of questionnaire points for each component are listed in Table 1.

As shown in Table 1, the first component positively correlated with scores for human-like, cute, friendly, warm, biological, conscious, natural, and emotional (correlation coefficients >.25). The second component positively correlated with intelligent, ethical, emotionally stable, and rational (correlation coefficients >.25), and negatively correlated with simple (correlation coefficients <−.25). Finally, the third component positively correlated with human-like, intelligent, and biological (correlation coefficients >.25), but negatively correlated with cute, friendly, active, positive, and curious (correlation coefficients <−.25). These results strongly indicate that participants formed different impressions of each opponent during the pre-scanning interaction session.

3.2. Mental function score

When we calculated mental function scores based on the results of the questionnaire, the mental function score

increased in the order of computer, Infanoid, Keepon, Actroid F, and human, indicating that participants tended to attribute mental functions to non-human opponents in this order.

3.3. Entropy

Fig. 3 shows the grand means of entropy for the five opponents across participants. A one-way ANOVA revealed a significant main effect of opponents [$F(4,15) = 6.40$, $p < .001$], and post-hoc t-tests revealed that entropies for human, Actroid F, and the computer were significantly higher than those for Infanoid and Keepon (Ryan's method, $p < .05$). This means that randomness in the series of left or right choices increased when the opponent was human, Actroid F, or a computer.

3.4. Relationship between PCA components and other behavioral measurements

When we evaluated correlations between PCA components and mental function scores and between PCA components and entropy, we found that the first PCA component (Table 1) was significantly correlated with the mental function score ($p < .05$, one-sample t-test across participants) but not with the entropy ($p < .05$, one-sample t-test across participants) (Fig. 4). In contrast, the third PCA component was significantly correlated with entropy ($p < .05$, one-sample t-test across participants) but not with mental function ($p < .05$, one-sample t-test across participants). No significant correlations were observed between the second PCA component and both behavioral variables. Thus, the first PCA component corresponded well to how much the participants attributed mental functions to the various opponents, but did not reflect that participants' mental factors influenced changing their game tactics. In contrast, the third PCA component well described participant's mental factors but not the attribution of mental function. Importantly, these two PCA components also had large loads of human-like in the impression questionnaire (Table 1).
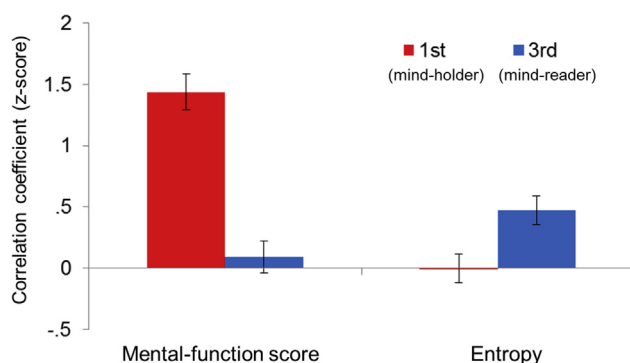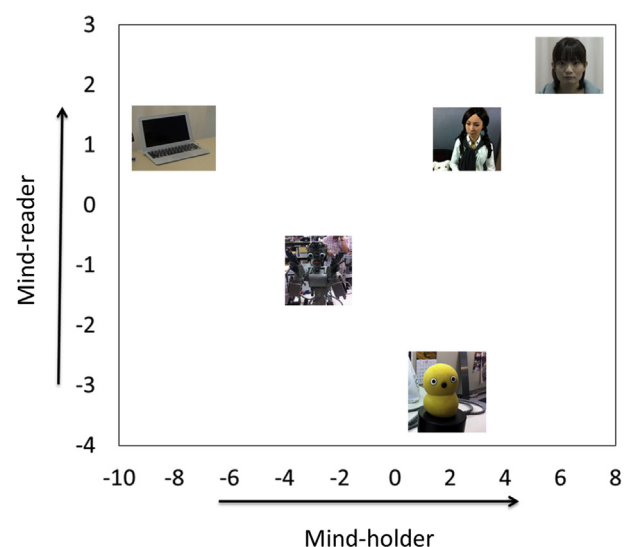
Based on these findings, we plotted specific values for each opponent obtained from the first and third PCA components in the x- and y-axes, respectively (Fig. 5). We found that the values increased in the order of computer, Infanoid, Keepon, Actroid F, and human along the x-axis, whereas the values became greater in the order of Keepon, Infanoid, computer, Actroid F, and human along the y-axis.

As we found that the first PCA component reflected the participants' attributions of mental function to the opponents and that the third PCA component corresponded to the participants' mental factors that led to changing game tactics in order to prevent their tactics being envisioned by their opponents (Fig. 4), we defined the x-axis as representing "mind-holderness" and the y-axis as representing "mind-readerness" of the opponents (Fig. 5). In light of this view, we could better explain that the Keepon was perceived by the participants as an agent with relatively high mind-holderness but less mind-readerness, in contrast an intelligent-looking computer was perceived as an agent with relatively high mind-readerness but less mind-holderness. Human opponents were classified as agents with higher mind-readerness and mind-holderness, and Actroid F (human-like android) was classified similarly. These results clearly indicate the multi-dimensionality in the perception of the various agents, which generally fits with a previous report (Gray et al., 2007).

3.5. Modulation of brain activity by preformed impression of an opponent

When we performed parametric modulation analysis, we found that activities in the bilateral mPFCs, posterior PCCs, TPJ/pSTS, and the left hippocampus were positively correlated with the regressor generated from the first PCA component (red areas in Fig. 6a and b). This suggests that the perceived mind-holderness of the opponents significantly modulated

Fig. 4 — Relationships between PCA components and mental function score (left) and between PCA components and entropy (right). The correlation coefficients were transformed to z-scores. The first PCA component was significantly correlated with the mental function score, whereas the third PCA component was significantly correlated with entropy.

Fig. 5 — Location of each opponent in two-dimensional space. The x-axis indicates "mind-holderness" and the y-axis indicates "mind-readerness" (see text). The score of PCA components for each opponent represents the mean value among participants.
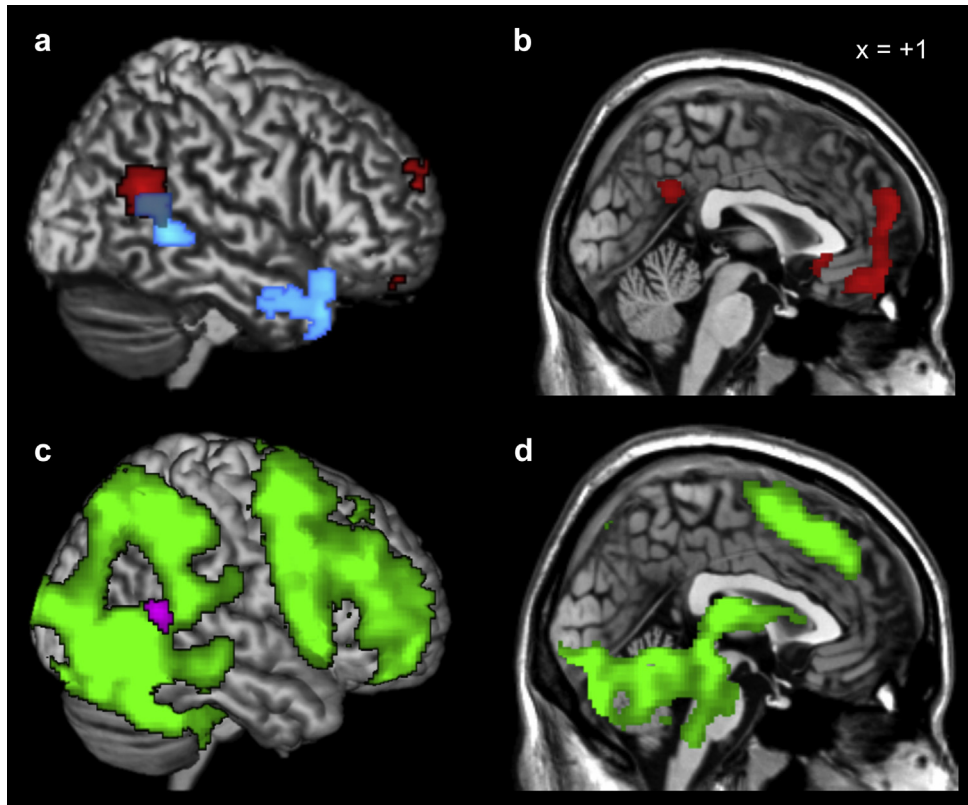
Fig. 6 — fMRI results. Brain regions where activities were modulated by "mind-holderness" (red) and "mind-readerness" (blue) are shown in panels (a) and (b). In panel (a), regions are superimposed on a lateral view of the MNI standard brain. In panel (b), regions are superimposed on a sagittal section, $x = +1$. In panels (c) and (d), regions activated during the game are shown in the corresponding images. The purple section in panel (c) represents a TPJ section where activity was modulated both by "mind-holderness" and by "mind-readerness", which also corresponded to the region activated during game playing.

the activities in these regions. On the other hand, the activities in the right TPJ/STS and TP were positively correlated with the regressor obtained from the third PCA component (blue areas in Fig. 6a). This indicates that the mind-readerness of the opponents significantly modulated the activities in these regions. The locations of voxels showing the strongest correlation with the regressors in each region are listed in Tables 2 and 3.

Importantly, activity in the same area of the right TPJ/pSTS was correlated both with first and third component regressors, but there was a tendency that the more posterior-dorsal aspect of this region was modulated by mind-holderness, whereas the anterior-ventral aspect was

modulated by mind-readerness. Taken together, two independent impressions of mind-holderness and mind-readerness obtained from the opponents through pre-scanning social interaction modulated the activity in distinct brain regions while participants played the game.

No areas that showed negative correlations with these two components were depicted. Likewise, no significant modulation of brain activity was observed in association with the second PCA component.

Finally, when we examined game-related brain regions, we found that several cortical and subcortical regions were activated during game playing (green areas in Fig. 6c and d). However, except for a small section in the TPJ (purple area in Fig. 6a), the above-mentioned regions did not exhibit significant activity increases during gameplay.

Table 2 — Regions where activities are modulated by mind-holderness.

Location	MNI coordinate			Z value	Cluster size (voxels)
	x	y	z		
mPFC	6	60	20	4.27	1793
Right TPJ/pSTS	46	−56	20	4.22	470
Left hippocampus	−14	−30	−4	3.95	273
Left TPJ/pSTS	−40	−58	24	3.87	291
Precuneus/PCC	−8	−52	36	3.79	416

Table 3 — Regions where activities are modulated by mind-readerness.

Location	MNI coordinate			Z value	Cluster size (voxels)
	x	y	z		
Right TPJ/pSTS	58	−46	0	4.63	293
Right TP	36	20	−32	4.33	638

4. Discussion

4.1. Impression affected later game tactics

In the present study, participants consistently played the game with a fixed computer algorithm pre-prepared by an experimenter. However, participants were instructed that they would play the game with each opponent, which would be indicated as a small icon at the top of the monitor inside the scanner. Indeed, post-scanning we confirmed that all of the participants believed that they played with five distinct opponents. Thus, changes in game tactics corresponding to each opponent largely depended on impressions that were formed during the previous interactive experiences (chatting).

This claim is supported by the following lines of evidence. First, the icon was too small for the participants to obtain online information about subtle changes in the opponents' facial expressions and behaviors. Thus, participants had to rely on their initial impressions about each opponent. Second, the entropy values obtained when the participants played with the computer were comparable with those obtained when playing against a human (Fig. 3). Interestingly, in a previous study where people played with a computer that did not show a lively flow of complex program code, the former was smaller than the latter (Takahashi et al., 2013). This indicates that impressions formed in the current study were obtained through pre-scanning interactions, and not from general preconceived notions. Finally, we found that hippocampal activity was modulated by mind-holderness (Table 2). It has been shown that people use their own repertoire of memories to predict the mental states of others, especially when the agents are similar to themselves (Perry, Hendler, & Shamay-Tsoory, 2011). Thus, the present hippocampal activity modulated by mind-holderness likely reflects a similar mental process where participants accessed their memories about the impressions they had made of each opponent.

Taken together, the previously obtained impression about opponents gained through direct interaction before the scanning likely affected game tactics that the participants employed in the scanner.

4.2. Mind-holderness and mind-readerness

In the present study, both mind-holderness and mind-readerness are defined in perceptual dimension. These are impressions of opponents received by the participants. In this sense, the present definition of mind-holderness and mind-readerness was not based on any particular substantial mental faculties of the opponents, which were not evaluated in the present study.

We found that the first PCA component reflected the degree to which participants attributed mental function to the opponents (Fig. 4), and that PCA values increased in the order of computer, Infanoid, Keepon, Actroid F, and human (Fig. 5). Thus, the horizontal mental axis in Fig. 5 reflects the opponent's degree of anthropomorphism, i.e., how much the participants thought that each opponent had mental function. Hence, the different degrees of perceived mind-holderness of the opponent likely affected the participant's mental

operations while they played the game, as indicated by its specific modulation of brain activity in a particular set of brain regions (see Fig. 6 and below).

The vertical mental axis in Fig. 5 (the third PCA component) should reflect different aspects of the opponent's characteristics, as this component was not correlated with the mental function score but correlated with behavioral entropy (Fig. 4). Greater entropy was observed when participants played with the computer, Actroid F, and human as compared with the Keepon and Infanoid (Fig. 2). It is theoretically known that, when people play this type of game, they tend to increase entropy in order to prevent their tactics from being read by an opponent (Nash, 1950). Based on this notion, together with our present entropy finding (Fig. 4), mind-readerness seems to be suitable to represent this metal axis. As described above, the present mind-readerness does not require any substantial ability of "mind reading" of an opponent, but reflect participants' impression that an opponent likely envisions their game tactics. In light of this view, it is worth discussing an opponent's possible gaze. We know from our previous study that when people play this game with a humanoid robot, behavioral entropy significantly increases in those who are sensitive to a robot's gaze compared to those who are not (Takahashi et al., 2013). This generally indicates that sensitivity to an opponent's gaze may increase behavioral entropy during the game, and the underlying mental states of participants are most likely strategizing to prevent their game tactics from being envisioned by the opponent. Thus, the present increase in participant entropy during gameplay with an intelligent-looking opponent with greater mind-readerness (computer, Actroid F, and human) indicates the existence of this type of cautious mental state.

One caveat to our conclusion is that there could be multiple alternative labels of the two mental axes (e.g., cool/warm, intelligence/emotion). An interesting avenue for future study is to refine the interpretations. On the other hand, it is worth noting that in the present study we demonstrated the multidimensionality of human social perceptions about other people/robots/agents and uncovered the underlying neural mechanisms.

4.3. Roles of two distinct sets of brain regions

4.3.1. Resemblance of social brain network and default mode network (DMN)

In the present study, we identified two distinct sets of brain regions with activities that were differently modulated depending on the perceived degrees of mind-holderness and mind-readerness of the opponents (Fig. 6). However, except for a small section in the TPJ (purple area in Fig. 6c), these sets of brain regions did not correspond to those with significantly increased activity during gameplay.

Among the modulated brain regions, the mPFC, PCC, and parts of the TPJ seem to correspond to brain regions that form the DMN, which shows increased activity during passive or resting periods compared to task periods (Raichle et al., 2001; Shulman et al., 1997). Thus, our findings are consistent with the accumulating evidence that the human social brain network closely resembles the DMN (Mars, Neubert, et al., 2012). Activities of these brain regions likely reflect ongoing

mental processes of broad- and unconstrained thoughts about the participant's bodily, perceptual, mental, and emotional states, which people call "random episodic silent thinking" (Andreasen et al., 1995) and "mind wandering" (Mason et al., 2007). Thus, these brain regions are essentially destined to deactivate during a "task" that requires participants to focus on achieving a specific purpose as explicitly constrained by an experimenter. In this vein, gameplay in the current study was our "task" that activated a wide range of cortical and subcortical brain regions (Fig. 6). On the other hand, unconstrained implicit thoughts that pertained to this task, e.g., active reading of the opponent's mental state and tactics to win a game, likely modulated the activity in the sets of brain regions.

Hereinafter, we argue two distinct roles of these sets of brain regions, which were uniquely modulated by the perception of the mind-holderness and mind-readerness attributed to opponents.

4.3.2. Mind-holderness

In the current study, we found that mind-holderness modulated activities in the posterior-dorsal section of the TPJ and in the networks of the PCC and mPFC (Fig. 6) that belong to the cingulum fiber tract (Thiebaut de Schotten, Dell'Acqua, Valabregue, & Catani, 2012). Humans possess the ability to make inferences about other people's mental states, such as the intentions and desires of others, and to refer to them to predict and explain behavior. This behavior, called mentalizing, is known to engage the social brain network (see Introduction), especially brain regions that belong to the cingulum tract (Amodio & Frith, 2006; Frith & Frith, 1999, 2003, 2006). The fact that the present brain regions directly corresponded to this mentalizing network indicates that the participants tried to mentalize an opponent's intention, tactics, and emotion during the game. Importantly, as the actual opponent was always the same computer algorithm that consistently used a fixed tactic irrespective of the opponent shown in the icon, this mentalizing was conducted by the participants based upon their formed pre-scanning social impressions of their opponents. Our claim that mentalizing occurs even for the present non-human opponents is corroborated by the finding that when people play a game with non-human agents, activities in the mentalizing network (mPFC and TPJ/STS) change according to the human-likeness of an agent (Krach et al., 2008). Taken together, the present activity modulation observed in this set of brain regions likely reflects the participants' mentalizing processes employed to read opponents' mental states, even for the non-human agents, depending on the degrees of opponents' mind-holderness.

4.3.3. Mind-readerness

In contrast, mind-readerness modulated activities in the anterior-ventral section of the TPJ including the pSTS and in the TP that belongs to the uncinate fascicule (UF) fiber tract (Thiebaut de Schotten et al., 2012). The UF tract is the largest tract of the fronto-temporal connections and is a ventral limbic pathway that originates rostrally in the temporal lobe and terminates in the ventral, medial, and orbital parts of the frontal cortex. This tract connects cortical regions involved in

visual and auditory recognition (superior and inferior temporal gyri) and in recognition memory (entorhinal, perirhinal, and parahippocampal cortices) with frontal areas implicated in emotion, inhibition, and self-regulation (Price et al., 2008; Schmahmann et al., 2007). Thus, the UF tract plays an important role in the interaction between cognition and emotion (Barbas, 2000; MacLean, 1952). In particular, the TP in the UF tract seems to be key in linking these two systems (Olson et al., 2007), which is necessary for complex information processing required for social interaction. Indeed, lesions involving the human UF tract may result in antisocial behaviors, probably due to the loss of self-regulation (Price et al., 2008). As described above, participants increased the randomness in their left or right choices (indicating entropy) depending on their perceived mind-readerness of opponents. Additionally, their underlying mental states were likely being employed for preventing their game tactics from being envisioned by intelligent-looking opponents. Considering the TP's function in linking cognition and emotion, the present activity modulation in the TP might be reflective of the participant's cautious mental states, which could have also affected their emotional states during the game.

4.3.4. The distinct functions of posterior-dorsal and anterior-ventral TPJ

We should also carefully discuss the finding that activities in different TPJ portions were uniquely modulated by mind-holderness and mind-readerness. Recently, it was shown that the TPJ can be subdivided into posterior-dorsal and anterior-ventral portions on the basis of its structural and functional connectivity (Mars, Neubert, et al., 2012; Mars, Sallet, et al., 2012). This finding indicates that these two TPJ areas play distinct roles, as has been previously suggested (Saxe, 2006).

The peak coordinate of the present TPJ region where activity was modulated by mind-holderness (Table 2) well-corresponded to those reported in many previous mentalizing tasks (Gallagher et al., 2000; Saxe & Kanwisher, 2003; Van Overwalle & Baetens, 2009). For example, one recent study has revealed that a person who has a strong tendency to attribute anthropomorphism (human characteristics) to animals or nonliving stimuli has greater gray matter volume in the TPJ (Cullen, Kanai, Bahrami, & Rees, 2013). Thus, the present activity modulation in the posterior-dorsal portion of the TPJ likely reflects different degrees of mentalizing that depended on how much participants anthropomorphized their opponents. Moreover, we assume that the role of the TPJ activity observed in this study was due to participants considering other's perspectives in order to interpret their opponents' internal states such as intention, emotion, and preference. This claim also seems to be supported by many previous findings (Blanke & Arzy, 2005; Decety, 2005; Jackson, Brunet, Meltzoff, & Decety, 2006; Ruby & Decety, 2004; Seger, Stone, & Keenan, 2004).

In contrast, the major role of the anterior-ventral TPJ/pSTS is social processing of an individual's physical signs especially as elicited from the face and eyes (Allison, Puce, & McCarthy, 2000; Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Van Overwalle & Baetens, 2009; Wicker, Michel, Henaff, & Decety, 1998). Based on our previous

finding that sensitiveness to a robot's gaze increases behavioral entropy during gameplay (Takahashi et al., 2013), we speculate that detecting possible gaze from opponents could be an important factor in modulating activity in this brain region. As the computer opponent does not have eyes, we assume that this region is also capable of detecting potential gaze not only from physical eyes, but also from the mind's eye.

4.4. Conclusion

Taken all together, when the opponent was an anthropomorphic mind-holder, participants took opponents' perspectives into account in order to mentalize their intention, tactics, and even emotion by recruiting the dorso-medial cingulum network. On the other hand, when the opponent was categorized as a mind-reader, participants became mindful of the possible gaze of the opponent, which could be reflected as modulation of activity in the anterior-ventral TPJ/pSTS. These results suggest that social interaction with mind-holder or mind-reader may distinctly shape the internal representation of our social brain, which may in turn determine how we behave for various agents that we encounter in our society.

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cortex.2014.03.011.

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences, 4*(7), 267–278.

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

Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G. L., et al. (1995). Remembering the past: two facets of episodic memory explored with positron emission tomography. *The American Journal of Psychiatry, 152*, 1576–1585.

Barbas, H. (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Research Bulletin, 52*(5), 319–330.

Blanke, O., & Arzy, S. (2005). The out-of-body experience: disturbed self-processing at the temporo-parietal junction. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 11*(1), 16–24.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction (roundtable series in behavioral economics).* Princeton, NJ: Princeton Univ. Press.

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., et al. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience, 6*, 103.

Cullen, H., Kanai, R., Bahrami, B., & Rees, G. (2013). *Social Cognitive and Affective Neuroscience.* Advance Access published July 24, 2013. 44, 0.

Decety, J. (2005). Perspective taking as the royal avenue to empathy. In B. F. Malle, & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 135–149). New York: Guilford Publications.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8*(11), 1611–1618.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: an experiment. *Etiology and Sociobiology, 14*, 247–256.

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

Frith, C. D. (1999). Interacting minds – a biological basis. *Science, 286*(5445), 1692–1695.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358*(1431), 459–473.

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

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

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*(1), 11–21.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*(3), 814–821.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619.

Haslam, N. (2006). Dehumanization: an integrative review. *Personality and Social Psychology Review, 10*(3), 252–264.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*(1), 80–84.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*(5), 752–761.

Kanda, T., Ishiguro, H., Ono, T., Imai, M., & Nakatsu, R. (2002). An evaluation on interaction between humans and an autonomous robot Robovie. *Journal of Robotics Society of Japan, 20*(3), 1–9.

Kozima, H. (2002). Infanoid: a babybot that explores the social environment. In K. Dautenhahn, A. Bond, L. Canamero, & B. Edmonds (Eds.), *Socially intelligent agents: Creating relationships with computers and robots.* Amsterdam: Kluwer Academic Publishers.

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2008). Keepon: a playful robot for research, therapy, and entertainment. *International Journal of Social Robotics, 1*(1), 3–18.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One, 3*(7), e2597.

Kuzmanovic, B., Bente, G., Von Cramon, D. Y., Schilbach, L., Tittgemeyer, M., & Vogeley, K. (2012). Imaging first impressions: distinct neural processing of verbal and nonverbal social information. *NeuroImage, 60*(1), 179—188.

Lisetti, C. L., Brown, S. M., Alvarez, K., & Marpaung, A. H. (2004). A social informatics approach to human-robot interaction with a service social robot. *IEEE Transactions on Systems, Man and Cybernetics, 34*(2), 195—209.

Loughnan, S., & Haslam, N. (2007). Animals and androids implicit associations between social categories and nonhumans. *Psychological Science, 18*, 116—121.

MacLean, P. D. (1952). Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Clinical Neurophysiology, 4*(4), 407—418.

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in Human Neuroscience, 6*(6), 189.

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. S. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral Cortex (New York, N.Y.: 1991), 22*(8), 1894—1903.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science (New York, N.Y.), 315*(5810), 393—395.

Nash, J. F. (1950). Bargaining problem. *Econometrica, 18*(2), 155—162.

Ohira, H., Matsunaga, M., & Murakami, H. (2013). Neural mechanisms mediating association of sympathetic activity and exploration in decision-making. *Neuroscience, 246*, 362—374.

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain: A Journal of Neurology, 130*(Pt 7), 1718—1731.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior, 15*(2), 123—142.

Perry, D., Hendler, T., & Shamay-Tsoory, S. G. (2011). Projecting memories: the role of the hippocampus in emotional mentalizing. *NeuroImage, 54*(2), 1669—1676.

Price, G., Cercignani, M., Parker, G. J. M., Altmann, D. R., Barnes, T. R. E., Barker, G. J., et al. (2008). White matter tracts in first-episode psychosis: a DTI tractography study of the uncinate fasciculus. *NeuroImage, 39*(3), 949—955.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 18*(6), 2188—2199.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 676—682.

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*(4), 1694—1703.

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

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*(2), 235—239.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage, 19*(4), 1835—1842.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., De Crespigny, A. J., et al. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain: A Journal of Neurology, 130*(Pt 3), 630—653.

Seger, C. A., Stone, M., & Keenan, J. P. (2004). Cortical activations during judgments about the self and another person. *Neuropsychologia, 42*(9), 1168—1177.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: 11. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience, 9*(5), 648—663.

Takahashi, H., Saito, C., Okada, H., & Omori, T. (2013). An investigation of social factors related to online mentalizing in a human-robot competitive game. *Japanese Psychological Research, 55*(2), 144—153.

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., & Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 48*(1), 82—96.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage, 48*(3), 564—584.

Wicker, B., Michel, F., Henaff, M. A., & Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. *NeuroImage, 8*(2), 221—227.

Yoshikawa, M., Matsumoto, Y., Sumitani, M., & Ishiguro, H. (2011). Development of an android robot for psychological support in medical and welfare fields. In *2011 IEEE international conference on Robotics and Biomimetics* (pp. 2378—2383).