Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres
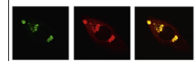
Research Report

Dissociable neural systems for moral judgment of anti- and pro-social lying

Akiko Hayashi[a,*], Nobuhito Abe[a], Toshikatsu Fujii[a], Ayahito Ito[a], Aya Ueno[a], Yuta Koseki[a], Shunji Mugikura[b], Shoki Takahashi[b], Etsuro Mori[a]

[a]Department of Behavioral Neurology and Cognitive Neuroscience, Tohoku University Graduate School of Medicine, Sendai, Japan
[b]Department of Diagnostic Radiology, Tohoku University Graduate School of Medicine, Sendai, Japan

ARTICLE INFO

ABSTRACT

Pro-social lying, which serves to benefit listeners, is considered more socially and morally acceptable than anti-social lying, which serves to harm listeners. However, it is still unclear whether the neural mechanisms underlying the moral judgment of pro-social lying differ from those underlying the moral judgment of anti-social lying. We used functional magnetic resonance imaging (fMRI) to examine the neural activities associated with moral judgment in anti- and pro-social lying. During fMRI scanning, subjects were provided with scenarios describing a protagonist's anti- and pro-social lying and were then asked to judge whether the protagonist's act was morally appropriate. The behavioral data showed that anti-social lying was mostly judged to be morally inappropriate and that pro-social lying was mainly judged to be morally appropriate. The functional imaging data revealed dissociable neural systems for moral judgment in anti- and pro-social lying. The anti-social lying, which was judged to be morally inappropriate, was associated with increased activity in the right ventromedial prefrontal cortex, right middle frontal gyrus, right precuneus/posterior cingulate gyrus, left posterior cingulate gyrus, and bilateral temporoparietal junction when compared with the control condition. The pro-social lying, which was judged to be morally appropriate, was associated with increased activity in the right middle temporal gyrus, right supramarginal gyrus, and the left middle cingulate gyrus when compared with the control condition. No overlapping activity was observed during the moral judgment of anti- and pro-social lying. Our data suggest that cognitive and neural processes for the moral judgment of lying are modulated by whether the lie serves to harm or benefit listeners.

© 2014 Elsevier B.V. All rights reserved.

Corresponding author. Present address: Department of Psychiatry, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan. Fax: +81 52 744 2293.
  E-mail address: akikohayashi@med.nagoya-u.ac.jp (A. Hayashi).

1. Introduction

Moral judgments are defined as evaluations (good vs. bad) of the actions or character of a person that are made with respect to a set of virtues held by a culture or subculture (Haidt, 2001). Moral judgments are important decision-making processes because the evaluation of our own actions or the actions of others facilitates social interactions. Recent functional neuroimaging studies of healthy subjects have explored the neural mechanisms involved in moral judgment (Berthoz et al., 2006; Greene et al., 2001, 2004; Harenski and Hamann, 2006; Heekeren et al., 2005; Moll et al., 2002a, 2002b, 2003). These studies have extended our understanding of the neural substrates of moral judgment, emphasizing the consistent involvement of the regions responsible for emotion, cognitive control, and mentalizing (for review, see Moll and Schulkin, 2009). Neuropsychological studies have also provided evidence regarding the neural substrates of moral judgment. For example, it was reported that patients with damage to the ventromedial prefrontal cortex, including those with frontotemporal dementia, show utilitarian moral judgment (Ciaramelli et al., 2007; Koenigs et al., 2007; Mendez et al., 2005).

During human social interaction, there are many types of complex social behaviors that inevitably involve moral judgment. One such behavior is deception, a psychological process by which one individual deliberately attempts to convince another person to accept as true what the liar knows to be false, typically in order for the liar, or sometimes for others, to gain some type of benefit or to avoid loss (Abe, 2009, 2011). Although many previous neuroimaging studies have shown the involvement of the prefrontal cortex in deception (Abe, 2009, 2011; Christ et al., 2009; Gombos, 2006; Hughes et al., 2005, Spence, 2004; Spence et al., 2004; Spence and Kaylor-Hughes, 2008), research into how the brain evaluates lies is of equal importance to research into the neural basis of lie generation.

We previously used positron emission tomography (PET) to explore the neural mechanisms underlying the willingness to forgive another person's moral transgression involving deception (Hayashi et al., 2010). During scanning, subjects were asked to judge the forgivability of a perpetrator's moral transgression. These transgressions were presented in four types of scenarios composed of a combination of two factors: the attitude of the perpetrator (dishonest or honest) and the severity of the moral transgression (serious or minor). We found that relative to honest responses, a perpetrator's dishonest responses were associated with right ventromedial prefrontal activity, which may reflect the subjects' identification of the perpetrator's deception. Further analysis revealed that the left ventromedial prefrontal cortex showed a significant interaction between the attitude of the perpetrator and the severity of the transgression, indicating that the right and left ventromedial prefrontal cortices play important, but somewhat different, roles in the forgiveness of moral transgressions involving deception.

It should be noted that we used only anti-social lying as the experimental stimuli in our PET study. Deception is generally thought to be an immoral, anti-social act, but it is not always an anti-social behavior. Lying can be a pro-social behavior that is intended to be altruistic. Pro-social lying is thought to be motivated by the desire to make others feel better or to spare their feelings and to facilitate smooth social relationships (DePaulo and Bell, 1996; DePaulo and Kashy, 1998). Unlike anti-social lies, which are usually made for personal gain, pro-social lies have benefits for the listener. Therefore, pro-social lying is often considered socially and morally acceptable (e.g., Bussey, 1999; Lindskold and Han, 1986; Lindskold and Walters, 1983). These findings naturally raise the question of whether the neural basis of moral judgments of pro-social lying differs from that of anti-social lying.

A study reported by Wu et al. (2011) touched on the above-mentioned question. The researchers used functional magnetic resonance imaging (fMRI) to examine the neural basis of evaluations of both lying and truth-telling in different social contexts. During fMRI scanning, they provided subjects with stories describing either lying or truth-telling about a protagonist's anti-social acts or good deeds. Their results demonstrated a differentiation between lying and truth-telling and between different types of lying in a network of brain regions. These regions included the bilateral superior frontal gyrus, bilateral inferior parietal lobule, bilateral cuneus, right lingual gyrus, right precuneus, and left postcentral gyrus. The results suggest that the judgments of lying and truth-telling involving a third party might be based on rational processing rather than emotional processing. However, Wu et al. (2011) did not find significant activation of the ventromedial prefrontal cortex, which is inconsistent with the findings of our previous study (Hayashi et al., 2010). Furthermore, the protagonist's lying in their experimental stimuli did not necessarily serve to help the listener. Thus, the neural basis of the moral judgment of pro-social lying has yet to be clarified.

The present study was designed to determine whether the moral judgment of pro-social lying differs from that of anti-social lying and how the ventromedial prefrontal cortex contributes to the two types of moral judgment. As with previous neuroimaging studies on decision-making in social settings, we used a scenario method (Greene et al., 2001; Sharot et al., 2007; Takahashi et al., 2009) (Fig. 1). During fMRI scanning, subjects were presented with scenarios in which the protagonist told the truth or a lie. Specifically, the protagonist's lying served to either harm (i.e., anti-social lying) or help (i.e., pro-social lying) the listener. The subjects were asked to judge whether the protagonist's act was morally appropriate.

Based on the previous findings described above, we tested the following two competing hypotheses: (a) if the "common currency" of basic neural activity is the functioning of the ventromedial prefrontal cortex, then ventromedial prefrontal activity is associated with the moral judgments of both anti- and pro-social lying; alternatively, and (b) if the ventromedial prefrontal cortex is sensitive to anti-social lying, then ventromedial prefrontal activity is associated with the moral judgment of only anti-social lying and not with pro-social lying. The results from our previous PET study provide no clues to favor either of these hypotheses. In that study, we observed the activation of the ventromedial prefrontal cortex, but because all of the stimuli were associated with anti-social acts, it was unclear whether the activation was specific to
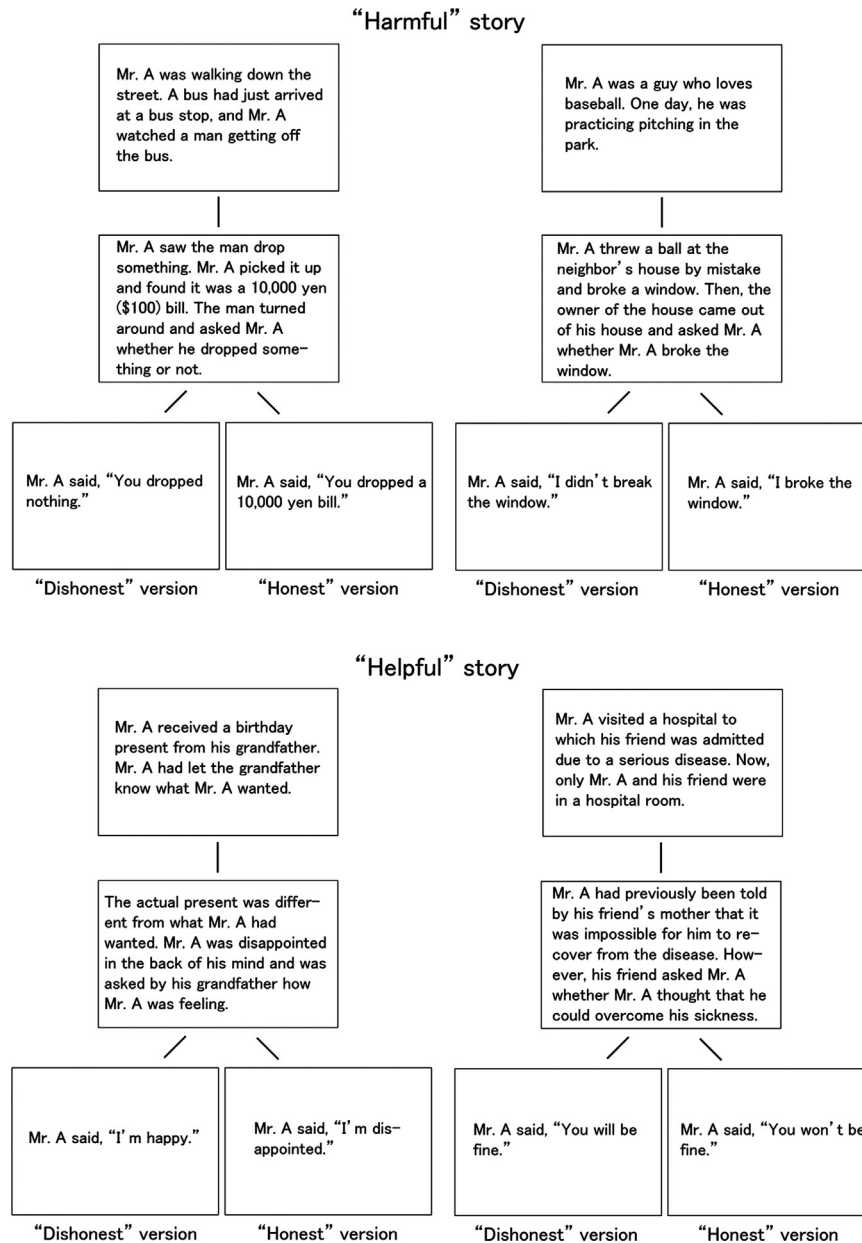
Fig. 1 – Two examples of the harmful and helpful stories used during the fMRI experiment. The harmful stories were those in which the protagonist's dishonesty would harm the listener and the protagonist would gain some type of benefit. The helpful stories were those in which the protagonist's dishonesty would help the listener and the listener would gain some type of benefit. Two different endings were prepared for each story, with dishonest and honest variations. The subjects were asked to judge whether the protagonist's actions in each story were morally appropriate.

anti-social lying. However, the latter hypothesis is supported by the theory that the judgment of whether a verbal statement is a lie is intrinsically a value judgment (e.g., Sweetser, 1987). In other words, whether an untruthful statement is a lie depends on whether it serves an anti- or pro-social purpose and whether it will harm or help the listener (see Harada et al., 2009). In line with this idea, Lee and Ross (1997) provided evidence that pro-social lying is judged to be less of a lie than anti-social lying. In this study, we provide evidence supporting this latter hypothesis, showing that neural systems for the moral judgment of anti- and pro-social lying are dissociable.

2.   Results

2.1.   Behavioral data

The mean proportion of responses and the reaction time of the subjects' responses for each experimental condition are shown in Table 1. For the proportion of "appropriate" responses in each experimental condition, we used a two-way repeated measures analysis of variance (ANOVA) with the story type (harmful or helpful) and the ending type (honest or dishonest) as factors. We found a near-significant main effect of the story type ($F[1, 36] = 3.289$,

Table 1 – The mean proportion and reaction time of the subjects' responses for each experimental condition.				
	Proportion (%)		Reaction time (ms)	
	Mean	SD	Mean	SD
Harmful/Dishonest				
"appropriate" response	18.2	13.4	2455	862
"inappropriate" response	81.8	13.4	1934	577
Harmful/Honest				
"appropriate" response	92.0	8.2	2081	641
"inappropriate" response	8.0	8.2	2617	1399
Helpful/Dishonest				
"appropriate" response	82.8	14.3	1928	511
"inappropriate" response	17.2	14.3	2149	982
Helpful/Honest				
"appropriate" response	32.3	17.8	2375	743
"inappropriate" response	67.7	17.8	2139	649

$p=0.078$), a significant main effect of the response type ($F[1, 36]=16.158$, $p<0.001$), and a significant interaction between the two factors ($F[1, 36]=613.291$, $p<0.001$). Post-hoc t-tests with Bonferroni corrections ($\alpha=0.05/4$) revealed significant differences between the Harmful/Dishonest and Harmful/Honest conditions ($p<0.001$), the Helpful/Dishonest and Helpful/Honest conditions ($p<0.001$), the Harmful/Dishonest and Helpful/Dishonest conditions ($p<0.001$), and the Harmful/Honest and Helpful/Honest conditions ($p<0.001$).

For the reaction time data, we used a two-way repeated measures ANOVA with the story type (harmful or helpful) and the ending type (honest or dishonest) as factors. Only the trials of "inappropriate" responses in the Harmful/Dishonest and Helpful/Honest conditions and "appropriate" responses in the Harmful/Honest and Helpful/Dishonest conditions were analyzed. We found a significant main effect of the ending type ($F[1, 36]=12.378$, $p<0.005$) but no main effect of the story type ($F[1, 36]=0.197$, $p=0.660$). The interaction between the two factors was not significant ($F[1, 36]=0.695$, $p=0.410$).

2.2. Imaging data

The neuroimaging results are summarized in Tables 2 and 3. First, to identify the neural systems involved in the moral judgment of anti-social lying, we compared the neural activity in the Harmful/Dishonest condition with the Harmful/Honest condition. This analysis revealed a significant difference in the blood oxygenation level-dependent (BOLD) signals of the right ventromedial prefrontal cortex, the right middle frontal gyrus, the right precuneus/posterior cingulate gyrus, the left posterior cingulate gyrus, and the bilateral temporoparietal junction (Fig. 2). The opposite comparison, the Harmful/Honest condition vs. the Harmful/Dishonest condition, revealed a significant difference in the BOLD signals of the right inferior occipital gyrus, the right lingual gyrus, and the left postcentral gyrus.

Second, to identify the neural systems involved in the moral judgment of pro-social lying, we compared neural activity in the Helpful/Dishonest condition with that in the Helpful/Honest condition. This analysis revealed a significant difference in the BOLD signals of the right supramarginal/postcentral gyrus, the right middle temporal gyrus/superior temporal sulcus, and the left middle cingulate gyrus (Fig. 3). The opposite comparison, the Helpful/Honest condition vs. the Helpful/Dishonest condition revealed a significant difference in the BOLD signals of the right putamen, the right calcarine cortex, the left superior frontal gyrus, the left insula, the left inferior occipital gyrus, the bilateral inferior frontal gyrus, and the bilateral middle temporal gyrus.

Third, to explicitly identify voxels where the effects were specific to the moral judgment of anti-social lying, we employed an exclusive masking procedure. The statistical parametric map constituting the exclusive mask was set at a threshold of $p<0.05$ (Helpful/Dishonest vs. Helpful/Honest), whereas the contrast to be masked was set at a threshold of $p<0.001$ (Harmful/Dishonest vs. Harmful/Honest). All of the regions showing even a very small effect in the comparison of Helpful/Dishonest vs. Helpful/Honest were thus excluded from the analysis of Harmful/Dishonest vs. Harmful/Honest. It should be noted that as the threshold of exclusive masking becomes more liberal, this masking procedure becomes more conservative. Thus, this analysis reveals the regions showing significant interaction effects (i.e., the regions showing a significant activity difference in the Harmful/Dishonest vs. Harmful/Honest comparison but not in the Helpful/Dishonest vs. Helpful/Honest comparison). We found a significant difference in the BOLD signals of the right ventromedial prefrontal cortex, the right middle frontal gyrus, the right precuneus/posterior cingulate gyrus, the left posterior cingulate gyrus, and the bilateral temporoparietal junction. These results indicate that the increased activities revealed by comparing the Harmful/Dishonest vs. Harmful/Honest conditions are highly specific to the moral judgment of anti-social lying.

Finally, to explicitly identify voxels where the effects were specific to the moral judgment of pro-social lying, we again employed an exclusive masking procedure. The statistical

Table 2 – Brain regions showing significant activation in each comparison.

Region (Brodmann's area)	Coordinates			Z value	Cluster size
	x	y	z		
Harmful/Dishonest vs. Harmful/Honest					
Right ventromedial prefrontal cortex (11)	6	35	−8	3.61	26
Right middle frontal gyrus (8)	27	20	52	4.20	100
Right precuneus/posterior cingulate gyrus (31)	3	−58	34	3.48	23
Right temporoparietal junction (39)	45	−64	25	4.42	119
Left posterior cingulate gyrus (31/23)	−9	−55	10	3.82	27
Left temporoparietal junction/middle occipital gyrus (19/39)	−39	−76	37	4.54	41
Harmful/Honest vs. Harmful/Dishonest					
Right inferior occipital gyrus (19)	36	−73	−8	3.80	32
Right lingual gyrus (18)	9	−85	−5	4.84	517
Left postcentral gyrus (4)	−54	−16	52	4.02	69

Table 3 – Brain regions showing significant activation in each comparison.

Region (Brodmann's Area)	Coordinates			Z value	Cluster size
	x	y	z		
Helpful/Dishonest vs. Helpful/Honest					
Right supramarginal gyrus/postcentral gyrus (40/2)	63	−25	37	4.23	142
Right middle temporal gyrus/superior temporal sulcus (37)	42	−61	10	3.57	27
Left middle cingulate gyrus (23)	−9	−28	46	3.96	51
Helpful/Honest vs. Helpful/Dishonest					
Right inferior frontal gyrus (45)	57	26	25	3.81	41
Right inferior frontal gyrus (45)	54	26	1	3.78	23
Right putamen	30	11	−8	4.47	107
Right middle temporal gyrus (21)	48	−25	−5	4.48	55
Right calcarine cortex (18)	18	−88	4	4.23	239
Left superior frontal gyrus (9)	−6	59	34	4.77	602
Left inferior frontal gyrus (45)	−48	29	7	3.83	64
Left insula	−30	17	−14	4.87	40
Left inferior frontal gyrus (44)	−39	11	25	4.16	60
Left middle temporal gyrus (21)	−57	2	−14	4.20	110
Left middle temporal gyrus (22)	−63	−19	−5	4.03	42
Left inferior occipital gyrus (19)	−33	−88	−8	4.32	202

parametric map constituting the exclusive mask had a threshold of $p < 0.05$ (Harmful/Dishonest vs. Harmful/Honest), and the contrast to be masked was a threshold of $p < 0.001$ (Helpful/Dishonest vs. Helpful/Honest). All of the regions showing even a very small effect in the comparison of Harmful/Dishonest vs. Harmful/Honest were thus excluded from the analysis of Helpful/Dishonest vs. Helpful/Honest. This analysis reveals the regions showing significant interaction effects (i.e., the regions showing a significant activity difference in the Helpful/Dishonest vs. Helpful/Honest comparison but not in the Harmful/Dishonest vs. Harmful/Honest comparison). We found a significant difference in the BOLD signals of the right supramarginal/postcentral gyrus, the right middle temporal gyrus/superior temporal sulcus, and the left middle cingulate gyrus. These results indicate that the increased activities revealed in the Helpful/Dishonest vs. Helpful/Honest comparison are highly specific to the moral judgment of pro-social lying.

3. Discussion

We used fMRI to examine whether the neural mechanisms underlying the moral judgment of anti-social lying differ from those generated during the moral judgment of pro-social lying. The functional imaging data revealed that anti-social lying that was judged to be morally inappropriate was associated with increased activity in the right ventromedial prefrontal cortex, the right middle frontal gyrus, the right precuneus/posterior cingulate gyrus, the left posterior cingulate gyrus, and the bilateral temporoparietal junction when compared with the control condition. The results also revealed that pro-social lying that was judged to be morally appropriate was associated with increased activity in the right middle temporal gyrus, the right supramarginal gyrus, and the left middle cingulate gyrus when compared with the control condition. No overlapping activity was observed
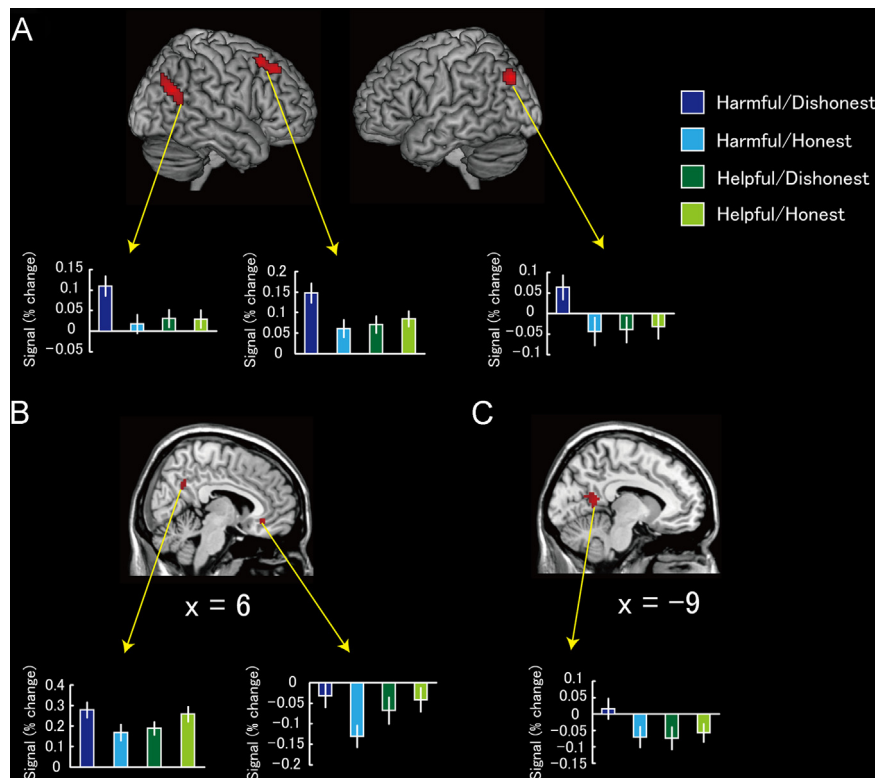
Fig. 2 – Significantly activated brain regions associated with moral judgment about anti-social lying, as revealed by a comparison of the Harmful/Dishonest condition with the Harmful/Honest condition. (A) The activations in the right middle frontal gyrus and the bilateral temporoparietal junction. (B) The activations in the right ventromedial prefrontal cortex and the right precuneus/posterior cingulate gyrus. (C) The activation in the left posterior cingulate gyrus.
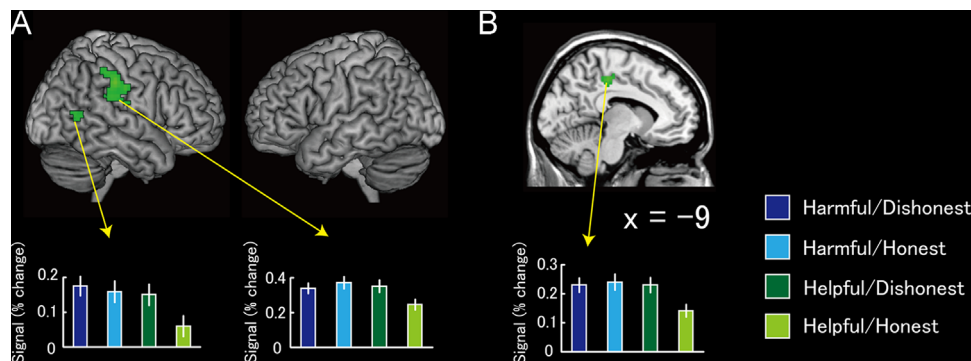
Fig. 3 – Significantly activated brain regions associated with moral judgment about pro-social lying, as revealed by a comparison of the Helpful/Dishonest condition with the Helpful/Honest condition. (A) The activation in the right supramarginal/postcentral gyrus and the right middle temporal gyrus/superior temporal sulcus. (B) The activation in the posterior part of the left middle cingulate gyrus.

during the moral judgment of anti- and pro-social lying. These results indicate that the moral judgment of lying is modulated by the protagonist's purpose and the social context. To our knowledge, this study is the first to provide evidence that distinct brain regions are engaged during the moral judgment of anti- and pro-social lying.

Our behavioral data confirmed that anti-social lying was mainly judged to be a morally inappropriate act and that pro-social lying was mainly judged to be a morally appropriate act. This interaction effect clearly indicates that the moral

judgment of lying depends on whether the lying serves to harm or help the listener. These results are consistent with previous studies showing that pro-social lying is often considered socially and morally acceptable (e.g., Bussey, 1999; Lindskold and Han, 1986; Lindskold and Walters, 1983). Another notable feature of our behavioral data is the reaction time result. We found a significant main effect of the ending type; that is, "dishonest" conditions required a shorter reaction time compared with "honest" conditions, regardless of whether the lying served to harm or help the target. These

results indicate that the processes of rejecting anti-social lying and accepting pro-social lying are default responses that may require shorter reaction times.

Consistent with our *a priori* hypothesis and our previous study (Hayashi et al., 2010), the right ventromedial prefrontal cortex was engaged during the moral judgment of anti-social lying, which was mainly judged to be morally inappropriate. We speculate that right ventromedial prefrontal activity is specifically associated with understanding the "malicious" intention of the protagonist and judging the "moral value" of lying. However, it should be noted that this region was deactivated in the Harmful/Honest condition and less deactivated in the Harmful/Dishonest condition. The ventromedial prefrontal cortex is known to be one of the regions associated with the default-mode network where task-induced deactivation is observed (Raichle et al., 2001; Watanabe, 2011). Thus, the present results are likely to be mainly derived from deactivation in the Harmful/Honest condition in which task-related cognitive demands might be increased. Alternatively, the smaller amount of deactivation in the right ventromedial prefrontal cortex during the Harmful/Dishonest condition may simply consist of both the increased activity for the process of understanding the harmful intention of the protagonist and the general task-related deactivation regardless of experimental conditions.

We found no significant difference in fMRI signals of the ventromedial prefrontal cortex between the moral judgment of pro-social lying, which was mainly judged to be morally appropriate, and its control condition. The specificity of the right ventromedial prefrontal cortex for the moral judgment of anti-social lying, and the fact that no overlapping activation was observed between moral judgments regarding two types of deception, jointly support the idea that different psychological processes underlie the moral judgment of pro- and anti-social lying. In a previous neuropsychological study, Stuss et al. (2001) showed that bilateral (particularly right) orbitofrontal lesions impair the ability to detect deception. It would be worth examining whether patients with ventromedial prefrontal lesions are able to understand the intention of a protagonist who tells a pro-social lie.

In addition to our hypothesized activity of the ventromedial prefrontal cortex, the results of the present study showed activations in the precuneus and the posterior cingulate gyrus. Notably, these regions, as well as the ventromedial prefrontal cortex, are located in the cortical midline structure, which is closely linked to the default-mode network. There have been reports of a relationship between the moral judgment and default-mode networks. For example, Harrison et al. (2008) showed that moral dilemma tasks were associated with increased activity in the default-mode network. Reniers et al. (2012) also observed increased activity in various parts of the networks for theory of mind, empathy, and default-mode when subjects performed moral decision-making tasks. According to these studies, the default-mode network plays an important role in moral judgment.

However, the default-mode network account does not explain why the cortical midline structures were specifically associated with the moral judgment of anti-social lying. Another possible explanation for the relationship between moral judgment and the cortical midline structures comes from accumulated findings on self-referential processing (for review, see Northoff et al., 2006). We speculate that self-referential processing plays an important role in the moral judgment of anti-social lying. In the harmful stories used in the present study, the protagonist harms the listener by telling lies. In most cases, such anti-social lies are told to gain some type of benefit for the liar. In this sense, anti-social lies, relative to pro-social lies, are more closely linked to the liars themselves (i.e., the self). Thus, the moral judgment of anti-social lying in harmful stories is more likely to require self-referential processing, when the subjects simulate the protagonist's action.

In addition to the cortical midline structures, we also observed activation in the temporal, parietal, and lateral frontal cortices during the moral judgment of anti-social lying. The activation of the temporoparietal junction in the current study may reflect the subject's reasoning that the protagonist's anti-social lying would cause harm to another. Consistent with this idea, Parkinson et al. (2011) found that the activation of the temporoparietal junction in response to the moral judgment of dishonest transgressions was associated with representing other people's beliefs. Young et al. (2007) also reported that the activation of the temporoparietal junction was associated with the moral judgment of the intent to harm others. As for the right middle frontal gyrus, it may seem reasonable to assume that the moral judgment of stories that consist of a protagonist's lie might require greater cognitive control than the moral judgment of stories without a protagonist's lie (e.g., Greene et al., 2001, 2004). However, this difference is unlikely given that the reaction time for moral judgment of anti-social lying is shorter than that for the control condition. Another possible interpretation is that this prefrontal activity reflects the successful resolution of conflict for moral judgments of anti-social lying. Once the context of the protagonist's lying is understood by the subject, the moral judgment of anti-social lying might become easier and less conflicting when compared with its control condition. The prefrontal activity might be responsible for the computation to resolve the conflict, thereby reducing the reaction time.

Pro-social lying that was judged to be morally appropriate was associated with the activation of several brain regions, including the right middle temporal gyrus extending to the superior temporal sulcus, the right supramarginal gyrus, and the posterior part of the left middle cingulate gyrus. Each of these regions might play different roles in social cognition that supports the moral judgment of pro-social behavior. For example, Fahrenfort et al. (2012) reported that individual differences in neural activity in the posterior superior temporal sulcus can predict pro-social behavior. Adolphs (2001) illustrated the component processes of social cognition, in which the right somatosensory area is responsible for the processing of social perception. Torta and Cauda (2011) showed that the middle/posterior cingulate gyrus is mainly involved in semantic and language processing.

Notably, the above-described activations associated with the moral judgment of pro-social lying overlapped with the regions implicated in the task-positive network (Fox et al., 2005). The task-positive network consists of brain regions activated during goal-directed task performance (Cabeza and

Nyberg, 2000). This network includes a set of brain regions called the "endogenous attention system" or the "dorsal attention system," which are active during directed attention (Corbetta and Shulman, 2002). Importantly, this task-positive network is known to be anticorrelated with the default-mode network (i.e., task-negative network) (Fox et al., 2005). Although the exact function of the task-positive network for the moral judgment of pro-social lying is obscure, these results provide further support for the idea that the cognitive and neural processes for the moral judgment of anti- and pro-social lying are highly dissociable.

In conclusion, the present study demonstrates that the neural systems for the moral judgment of lying are modulated by whether the lie serves to harm or help the listener. Specifically, the cortical midline structures are thought to be a core neural network for the moral judgment of anti-social lying, whereas the task-positive network is thought to play an important role in the moral judgment of pro-social lying. These results suggest that the moral judgment of pro-social lying is fundamentally different from the moral judgment of anti-social lying, despite the fact that these types of lying are both associated with protagonists' untruthful statements.

The present study has limitations. First, the scenarios used are fictitious, and the protagonist's actions in the scenarios do not produce any real harmful or helpful outcomes. In this regard, we acknowledge that there still may be a gap between the patterns of brain activity observed in laboratory settings and those observed in real-life situations. Second, although the harmful and helpful stories were matched for familiarity, emotional valence, imaginability, and the number of letters, the content of the stories was not sufficiently controlled. For example, some stories were associated with monetary gain/loss, but others were not, which might affect the present results. Nevertheless, we believe that our results further clarify the neural mechanisms underlying complex moral judgments in social settings.

4. Experimental procedures

4.1. Subjects

Forty volunteers (19 males and 21 females, age range 20–25 years, mean age 21.3 years) with no history of neurological or psychiatric disease were paid for their participation in this study. All the subjects were right-handed, as determined by the Edinburgh Handedness Inventory (Oldfield, 1971). Data from three subjects were excluded due to scanner malfunctions. Thus, our results are based on the remaining 37 volunteers (18 males and 19 females, age range 20–25 years, mean age 21.2 years). After receiving a detailed description of the study, all participants gave their written informed consent in accordance with the Declaration of Helsinki and the guidelines approved by the Ethical Committee of Tohoku University.

4.2. Stimuli and tasks

We prepared 180 stories, all of which described common, but fictitious, events that could occur in real-life situations. Each

story contained two characters, including a protagonist and another character (i.e., the person listening to the lying or truth-telling by the protagonist). First, we separated these stories into harmful and helpful groups. In the harmful stories, if the protagonist tells a lie to the listener, the lie harms the listener and the protagonist gains some type of benefit. In the helpful stories, if the protagonist tells a lie to the listener, the lie helps the listener and the listener gains some type of benefit. To validate the distinction between the harmful and helpful stories, we recruited a separate group of 16 subjects (8 males and 8 females, age range 18–29 years, mean age 20.8 years) who did not participate in the fMRI experiment. These subjects rated whether the lying in each story served to harm or help the listener on a 7-point scale. The scale ranged from completely harmful (7) to completely helpful (1). Based on the mean rating scores of each story, we chose 48 stories ranked in descending order (mean=5.31, SD=0.56) and 48 stories ranked in ascending order (mean=2.96, SD=0.55) for the fMRI experiment. The former 48 stories and the l the latter 48 stories were labeled as "harmful" and "helpful" stories, respectively. All 48 harmful stories scored greater than 4, and all 48 helpful stories scored less than 4. A t-test confirmed that there was a significant difference in the mean rating scores between these two sets of stories ($p < 0.05$). The two sets of harmful and helpful stories were also matched for familiarity, emotional valence, and imaginability (all p values > 0.1), as rated by the separate group of 16 subjects, using a 7-point scale for each criterion. The total number of letters for each part of each story was also matched between the harmful and helpful stories (all p values > 0.1).

We then added one sentence at the end of each story that described whether the protagonist told a lie or the truth. Consequently, the stimuli consisted of the following four types of stories: (a) 24 stories in which the protagonist harms the listener by telling a lie (i.e., dishonest acts in harmful stories; Harmful/Dishonest), (b) 24 stories in which the protagonist does not harm the listener by telling the truth (i.e., honest acts in harmful stories; Harmful/Honest), (c) 24 stories in which the protagonist helps the listener by telling a lie (i.e., dishonest acts in helpful stories; Helpful/Dishonest), and (d) 24 stories in which the protagonist does not help the listener by telling the truth (i.e., honest acts in helpful stories; Helpful/Honest). Each story was separated into three parts: the introduction, the event, and the protagonist's act. The introduction consisted of two sentences and described the background of the event. The event also consisted of two sentences and described a situation involving the protagonist and another person that required the protagonist to take action. The protagonist's act consisted of one sentence revealing whether the protagonist told a lie. Fig. 1 illustrates examples of these stories. During fMRI scanning, half of the subjects were presented with this set of 96 stimuli, and the remaining subjects were presented with another set of 96 stimuli in which the protagonist's act (lying or truth-telling) in each story was reversed.

During fMRI scanning, the subjects were presented with a total of 96 stories (48 harmful and 48 helpful stories) in a predetermined, pseudorandom order over three consecutive runs. The text of the stories was presented in white on the center of a black screen. As previously noted, each story

consisted of three parts (the introduction, the event, and the protagonist's act). For each story, the introduction was presented for 6 s and was followed by a fixation cross for 3 s. The event was then presented for 6 s and was followed by a fixation cross for 3 s. The protagonist's act was presented for 6 s and was followed by a fixation cross for 6 s. Each trial lasted a total of 30 s. After the subjects read the last part, they were asked to judge whether the protagonist's act was morally appropriate by pressing a button (right forefinger for "morally appropriate" and right middle finger for "morally inappropriate"). Although there was no time limit for judging the protagonist's act, the subjects were asked to make a moral judgment of the protagonist's act as soon as possible after reading the last part.

4.3.   Data acquisition and analysis

Whole-brain imaging was performed with a 3.0-Tesla MRI scanner (MAGNETOM Trio, A Tim System; Siemens-Asahi Medical Technologies Ltd., Tokyo, Japan). A T2∗-weighted echo planar imaging (EPI) sequence sensitive to BOLD contrast was used for functional imaging with the following parameters: repetition time (TR)=2500 ms, echo time (TE)= 30 ms, flip angle=90°, 80×80 acquisition matrix, field of view (FOV)=240 mm, and in-plane resolution=3 mm × 3 mm. Forty-three axial slices with a slice thickness of 3 mm and an interslice gap of 0.5 mm were obtained. We used a tilted acquisition sequence at 30° to the AC-PC line to recover the magnetic susceptibility-induced signal losses due to the sinus cavities (Deichmann et al., 2003). A high-resolution (spatial resolution $1 \times 1 \times 1$ mm^3) structural image was also acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) pulse sequence. Head motion was restricted using firm padding that surrounded the head. Visual stimuli were presented through magnet-compatible goggles. The subjects' responses were collected using a magnet-compatible response box. The EPI images were acquired in three consecutive runs. The first five scans in each run were discarded to allow for T1 equilibration effects.

Data preprocessing and statistical analyses were performed using the SPM8 software (Wellcome Department of Imaging Neuroscience, London, UK). All volumes acquired from each subject were corrected for different slice acquisition times. The resultant images were then realigned to correct for small movements occurring between scans. This process generated an aligned set of images and a mean image for each subject. Each participant's T1-weighted structural MRI was coregistered to the mean of the realigned EPI images and segmented to separate out the gray matter, which was normalized to the gray matter in a template image based on the Montreal Neurological Institute (MNI) reference brain. Using the parameters from this normalization process, the EPI images were also normalized to the MNI template (re-sampled voxel size 3 mm × 3 mm × 3 mm) and smoothed with an 8 mm full-width at half-maximum Gaussian kernel.

For each participant, an event-related analysis was first conducted on a voxel-by-voxel basis, in which all instances of a particular event type at the time of the participants' responses were modeled through convolution with a canonical hemodynamic response function. There were a total of four experimental conditions: Harmful/Dishonest, Harmful/Honest, Helpful/Dishonest, and Helpful/Honest. In the Harmful/Dishonest and Helpful/Honest conditions, the trials producing "appropriate" responses, which were thought to be a relatively small number of trials, were excluded from the analyses to avoid potential confounds between two types of responses. For a similar reason, in the Harmful/Honest and Helpful/Dishonest conditions, the trials with responses of "inappropriate" were excluded from the analyses. The parameter estimates (betas) for each condition were calculated for all brain voxels, and the relevant contrasts of parameter estimates were computed. A high-pass filter of 1/128 Hz was used to remove low-frequency noise, and an AR (1) model was used to correct for temporal autocorrelations. The contrast images for all subjects were then entered into a series of one-sample t-tests. This procedure permitted statistical inference at the population level (random-effects analysis). The significant activations were identified at a statistical threshold of $p < 0.001$ (uncorrected for multiple comparisons) with a cluster size of 20 or more voxels. The peak voxels of clusters exhibiting reliable effects are reported in the MNI coordinates. MarsBaR software was used to extract the percentage signal changes in activated or deactivated regions (Brett et al., 2002).

In addition to the standard subtraction analyses, we also employed an exclusive masking procedure to determine whether subjects activated dissociable neural networks during the moral judgment of anti- and pro-social lying. Exclusive masking was applied to the contrast of interest to eliminate voxels activated by the other contrast. Namely, the contrast of Harmful/Dishonest vs. Harmful/Honest was exclusively masked with the contrast of Helpful/Dishonest vs. Helpful/Honest and vice versa. The significance level of the masks was set at $p < 0.05$ (uncorrected for multiple comparisons).

Acknowledgments

R E F E R E N C E S

Abe, N., 2009. The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. Curr. Opin. Neurol. 22, 594–600.

Abe, N., 2011. How the brain shapes deception: an integrated review of the literature. Neuroscientist 17, 560–574.

Adolphs, R., 2001. The neurobiology of social cognition. Curr. Opin. Neurobiol. 11, 231–239.

Berthoz, S., Grezes, J., Armony, J.L., Passingham, R.E., Dolan, R.J., 2006. Affective response to one's own moral violations. Neuroimage 31, 945–950.

Brett, M., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox [abstract]. Neuroimage 16 (Supplement 1) (abstract 497).

Bussey, K., 1999. Children's categorization and evaluation of different types of lies and truths. Child Dev. 70, 1338–1347.

Cabeza, R., Nyberg, L., 2000. Imaging cognition II: an empirical review of 275 PET and fMRI studies. J. Cogn. Neurosci. 12, 1–47.

Christ, S.E., Van Essen, D.C., Watson, J.M., Brubaker, L.E., McDermott, K.B., 2009. The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. Cereb. Cortex 19, 1557–1566.

Ciaramelli, E., Muccioli, M., Ladavas, E., di Pellegrino, G., 2007. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. Soc. Cogn. Affect. Neurosci. 2, 84–92.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215.

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R., 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430–441.

DePaulo, B.M., Bell, K.L., 1996. Truth and investment: lies are told to those who care. J. Person. Soc. Psychol. 71, 703–716.

DePaulo, B.M., Kashy, D.A., 1998. Everyday lies in close and casual relationships. J. Personal. Soc. Psychol. 74, 63–79.

Fahrenfort, J.J., van Winden, F., Pelloux, B., Stallen, M., Ridderinkhof, K.R., 2012. Neural correlates of dynamically evolving interpersonal ties predict prosocial behavior. Front. Neurosci. 6, 28.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. USA 102, 9673–9678.

Gombos, V.A., 2006. The cognition of deception: the role of executive processes in producing lies. Genet. Soc. Gen. Psychol. Monogr. 132, 197–214.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. Neuron 44, 389–400.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293, 2105–2108.

Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. 108, 814–834.

Harada, T., Itakura, S., Xu, F., Lee, K., Nakashita, S., Saito, D.N., Sadato, N., 2009. Neural correlates of the judgment of lying: a functional magnetic resonance imaging study. Neurosci. Res. 63, 24–34.

Harenski, C.L., Hamann, S., 2006. Neural correlates of regulating negative emotions related to moral violations. Neuroimage 30, 313–324.

Harrison, B.J., Pujol, J., Lopez-Sola, M., Hernandez-Ribas, R., Deus, J., Ortiz, H., Soriano-Mas, C., Yucel, M., Pantelis, C., Cardoner, N., 2008. Consistency and functional specialization in the default mode brain network. Proc. Natl. Acad. Sci. USA 105, 9781–9786.

Hayashi, A., Abe, N., Ueno, A., Shigemune, Y., Mori, E., Tashiro, M., Fujii, T., 2010. Neural correlates of forgiveness for moral transgressions involving deception. Brain Res. 1332, 90–99.

Heekeren, H.R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H.P., Villringer, A., 2005. Influence of bodily harm on neural correlates of semantic and moral decision-making. Neuroimage 24, 887–897.

Hughes, C.J., Farrow, T.F., Hopwood, M.C., Pratt, A., Hunter, M.D., Spence, S.A., 2005. Recent developments in deception research. Curr. Psychiatry Rev. 1, 273–279.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A., 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446, 908–911.

Lee, K., Ross, H.J., 1997. The concept of lying in adolescents and young adults: testing Sweetser's folkloristic model. Merrill-Palmer Q. 43, 255–270.

Lindskold, S., Han, G., 1986. Intent and the judgment of lies. J. Soc. Psychol. 126, 129–130.

Lindskold, S., Walters, P.S., 1983. Categories for acceptability of lies. J. Soc. Psychol. 120, 129–136.

Mendez, M.F., Anderson, E., Shapira, J.S., 2005. An investigation of moral judgement in frontotemporal dementia. Cogn. Behav. Neurol. 18, 193–197.

Moll, J., de Oliveira-Souza, R., Bramati, I.E., Grafman, J., 2002a. Functional networks in emotional moral and nonmoral social judgments. Neuroimage 16, 696–703.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., 2003. Morals and the human brain: a working model. Neuroreport 14, 299–305.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., Bramati, I.E., Mourao-Miranda, J., Andreiuolo, P.A., Pessoa, L., 2002b. The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. J. Neurosci. 22, 2730–2736.

Moll, J., Schulkin, J., 2009. Social attachment and aversion in human moral cognition. Neurosci. Biobehav. Rev. 33, 456–465.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., Panksepp, J., 2006. Self-referential processing in our brain—a meta-analysis of imaging studies on the self. Neuroimage 31, 440–457.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P.E., Mendelovici, A., McGeer, V., Wheatley, T., 2011. Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. J. Cogn. Neurosci. 23, 3162–3180.

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. Proc. Natl. Acad. Sci. USA 98, 676–682.

Reniers, R.L., Corcoran, R., Vollm, B.A., Mashru, A., Howard, R., Liddle, P.F., 2012. Moral decision-making, ToM, empathy and the default mode network. Biol. Psychol. 90, 202–210.

Sharot, T., Riccardi, A.M., Raio, C.M., Phelps, E.A., 2007. Neural mechanisms mediating optimism bias. Nature 450, 102–105.

Spence, S.A., 2004. The deceptive brain. J. R. Soc. Med. 97, 6–9.

Spence, S.A., Hunter, M.D., Farrow, T.F., Green, R.D., Leung, D.H., Hughes, C.J., Ganesan, V., 2004. A cognitive neurobiological account of deception: evidence from functional neuroimaging. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 1755–1762.

Spence, S.A., Kaylor-Hughes, C.J., 2008. Looking for truth and finding lies: the prospects for a nascent neuroimaging of deception. Neurocase 14, 68–81.

Stuss, D.T., Gallup Jr., G.G., Alexander, M.P., 2001. The frontal lobes are necessary for 'theory of mind'. Brain 124, 279–286.

Sweetser, E.E., 1987. The definition of lie: an examination of the folk models underlying a semantic prototype. In: Holland, D. (Ed.), Cultural Models in Language and Thought. Cambridge University Press, New York, pp. 43–66.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., Okubo, Y., 2009. When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. Science 323, 937–939.

Torta, D.M., Cauda, F., 2011. Different functions in the cingulate cortex, a meta-analytic connectivity modeling study. Neuroimage 56, 2157–2172.

Watanabe, M., 2011. Are there internal thought processes in the monkey? Default brain activity in humans and nonhuman primates. Behav. Brain Res. 221, 295–303.

Wu, D., Loke, I.C., Xu, F., Lee, K., 2011. Neural correlates of evaluations of lying and truth-telling in different social contexts. Brain Res. 1389, 115–124.

Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. Proc. Natl. Acad. Sci. USA 104, 8235–8240.