doi:10.1093/scan/nst134 SCAN (2014) 9, 1472–1480

Detection of deception based on fMRI activation patterns underlying the production of a deceptive response and receiving feedback about the success of the deception after a mock murder crime

Qian Cui,[1,2] Eric J. Vanman,[3] Dongtao Wei,[1,2] Wenjing Yang,[1,2] Lei Jia,[1,2] and Qinglin Zhang[1,2]
[1]Faculty of Psychology, Southwest University, Chongqing, China, [2]Key Laboratory of Cognition and Personality (Southwest University), Ministry of Education, China, and [3]School of Psychology, University of Queensland, Queensland, Australia

The ability of a deceiver to track a victim's ongoing judgments about the truthfulness of the deceit can be critical for successful deception. However, no study has yet investigated the neural circuits underlying receiving a judgment about one's lie. To explore this issue, we used a modified Guilty Knowledge Test in a mock murder situation to simultaneously record the neural responses involved in producing deception and later when judgments of that deception were made. Producing deception recruited the bilateral inferior parietal lobules (IPLs), right ventral lateral prefrontal (VLPF) areas and right striatum, among which the activation of the right VLPF contributed mostly to diagnosing the identities of the participants, correctly diagnosing 81.25% of 'murderers' and 81.25% of 'innocents'. Moreover, the participant's response when their deception was successful uniquely recruited the right middle frontal gyrus, bilateral IPLs, bilateral orbitofrontal cortices, bilateral middle temporal gyrus and left cerebellum, among which the right IPL contributed mostly to diagnosing participants' identities, correctly diagnosing 93.75% of murderers and 87.5% of innocents. This study shows that neural activity associated with being a successful liar (or not) is a feasible indicator for detecting lies and may be more valid than neural activity associated with producing deception.

Keywords: deception; judgment; lie detection; functional magnetic resonance imaging

INTRODUCTION

In real life, deception is a highly complex social and cognitive process that involves decision making, risk taking, cognitive control, mentalizing and reward processing (Sip *et al.*, 2008). However, prior experimental researches have focused only on the cognitive control involved in the production of a deceptive response (Langleben *et al.*, 2002, 2005; Kozel *et al.*, 2004a,b, 2005, 2009a,b; Davatzikos *et al.*, 2005; Ganis *et al.*, 2011). Less attention has been given to the consequences of the deception—that is, when the interlocutor judges whether the deceptive behavior was truthful or deceptive, which is typically the deceiver's primary concern. Sip *et al.* (2012) have recently reported that whether the deceiver would be confronted about his responses by the interlocutor affects the neural circuits underlying deception production, thus highlighting the importance of the interlocutor's judgment in deception. No study, however, has directly investigated the neural circuits that underlie that judgment processing during the deception.

Rather than the cognitive and neural mechanisms underlying deception, researchers have focused on the more interesting and attractive issue of pursuing markers of deception and obtaining accurate determinations of veracity. Recent studies have used functional magnetic resonance imaging (fMRI) technology to identify the neural signals associated with deception, which have been subsequently used as markers to detect lies. Most of these studies have focused on the neural signals involved in deceptive responses, which primarily involve the activation of the executive system, including the frontal–parietal and anterior cingulate cortices, to infer deception (for reviews, see Wolpe *et al.*, 2005; Langleben, 2008; Sip *et al.*, 2008; Abe, 2011). This method has been criticized on the basis that drawing inferences about deception from the activation of the executive system is fraught with error because the executive system is merely involved in deception but not unique to it (Poldrack, 2006; Sip *et al.*, 2008; Abe, 2011). Given that deception is too complex to be measured by any single biological response, we may have more success if we use other indirect markers of deception instead of trying to detect it directly (Sip *et al.*, 2008). With this consideration, we investigated whether neural signals in response to the interlocutor's judgment in the deception context could be used as a reliable marker for inferring deception—a question that has not yet been investigated.

To answer this question, we designed a modified Guilty Knowledge Test (GKT) to record the brain activation patterns that are related to producing deceptive responses and processing judgments given by an interlocutor after a 'mock murder' situation. This modified GKT uses a standard three-kinds-of-stimulus design, which comprises probes, targets and irrelevants; however, the difference from the standard GKT is that after the participants respond to each item, the computer gives a judgment indicating whether the previous response was truthful or deceptive, allowing one to record the brain activation of the deceivers as they see how successful (or not) they were. We hypothesized that if the participants were (mock) murderers who responded deceptively to the probes, then the judgment following their deceptive response would result in greater brain activation because they cared so much about the outcome. That is, their brain activation patterns would be different from those participants who were innocent and responded truthfully to the probes, especially when the judgment indicated that the deceiver was wrongly judged to be truthful, which would evoke complex emotional and cognitive processes, such as secret delight, conflict recognition and error processing. Furthermore, for each of

Received 31 January 2013; Revised 18 July 2013; Accepted 9 August 2013
Advance Access publication 14 August 2013
Authors thank Zhiyi Sun and Ruihan Chen for their assistance with this study.
This study was supported by the National Natural Science Foundation of China (30970892, 31170983), http://www.nsfc.gov.cn/Portal0/default152.htm. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Correspondence should be addressed to Qinglin Zhang, Faculty of Psychology, Southwest University, Chongqing 400715, China. E-mail: zhangql@swu.edu.cn

these two stages of deception (i.e. producing deception and receiving the judgment of the interlocutor), we separately conducted discriminant analyses to examine the brain regions that contribute to diagnosing whether the participants were guilty or innocent; we computed their rates of accurate diagnosis using a cross-validation method similar to that used by *Nose et al.* (2009).

METHODS

Participants

A total of 32 undergraduates from a university in southwestern China participated in this study and were monetarily compensated after the experiment. Of 32 participants, 16 [$n = 8$ males, 8 females; mean age 20.94 (s.d. 1.24) years] were recruited at first to be 'murderers' (MUD group); the exact results about judgment processing will be reported in another paper in which we use a different analytical approach from that used in this study. One year later, the other 16 participants [$n = 9$ males, 7 females; mean age, 21.38 (s.d. 1.63) years] were recruited to be 'innocents' (INC group) and underwent the same paradigm as that of the 'murderers' 1 year before. This study used the data of all 32 participants. Every participant was right-handed and free from any physiological or psychological disease. After the procedures were fully explained, all of the participants signed an informed written consent according to the Declaration of Helsinki (Lynoe *et al.*, 1991). This study was approved by the local Ethics Committee.

Materials and procedure

The materials and procedures for the two groups were the same, including three phases: mock murder, lie detection and post-scan briefing report.

Mock murder

At first, each participant was instructed about the rules of a 'mock murder game' by the investigator (Z.Y.S.) and was introduced to three co-players, against whom they would play later and who were actually confederates. Subsequently, four players were asked to draw a card labeled either 'Murderer' or 'Innocent', which represented their identities during the game. Afterward, they were brought to four separate rooms. Actually, all of the participants in the MUD group were predetermined to be 'murderers', whereas all of the participants in the INC group were predetermined to be 'innocents'. All participants believed that one murderer existed among the four players and that the remaining three players were innocents. Furthermore, they believed that the identity of each player was known only by Z.Y.S. and that the player's identity was anonymous to any other person, including the three other co-players.

To 'commit a murder', each participant of the MUD group was required to complete a 'mock murder questionnaire' (Supplementary material A), including seven items that asked the murderer to (i) write down two names of his or her friends and choose one of the two friends to kill; and (ii) determine the time of the crime, (iii) the weapon, (iv) the part of body to attack, (v) the color of the sack in which to place the corpse and (vi) the place to conceal the corpse. The six selected crime details were later used as probes in the GKT. After completing the questionnaire, the murderer was asked to create a story as impressively as possible regarding the murder with these crime details and to imagine that he or she was actually committing such a murder. They were instructed to repeat the story sufficiently until Z.Y.S. was assured that the participants could remember all six crime details and that the story had become adequately concrete. Meanwhile, the participants of the INC group were asked only to complete a questionnaire regarding their demographic information. Subsequently, the participants were given instructions according to their identities (Supplementary material B) by Z.Y.S. and they were taken to the scanning room to take a lie detector test, which was conducted by another investigator (Q.C.).

Lie detection

Three types of words (Supplementary material C) were visually displayed in the lie detection phase: probes (P), the crime details determined by the murderer in the mock murder phase; targets (T), words that were irrelevant to the murder but were memorized by all of the participants before scanning and later required a unique response during scanning (we used randomly appearing targets to force participants to pay attention to the screen); and irrelevants (I), new words that had nothing to do with the murder. The probes, targets and irrelevants used for the murderers were the same as those used for the innocents. However, considering that the innocents could not distinguish probes from irrelevants, probes were equivalent to irrelevants for them. As one word was displayed on the screen, the participants indicated whether they had seen that word before. Afterward, the computer would pronounce a judgment. The participants were informed that, for the target, judgment would be decided according to whether they responded correctly. Specifically, if they responded correctly (yes) to targets, the judgment would be '+2' (TP, positive judgment following target) because they successfully recognized targets that they had memorized before scanning. In contrast, if they responded incorrectly (no) to targets, then the judgment would be '−2' (TN, negative judgment following target) because they could not recognize the memorized items or they did not concentrate on the experiment. For each of the other words, the computer would judge whether they told the truth or a lie based on their brain activities recorded by fMRI as they responded to each word and would subsequently give a judgment of '+2' or '−2,' respectively. Unknown to the participants, the judgments regarding the probes and irrelevants were experimentally predetermined by the experimenter; therefore, the judgments could equally be likely to be positive (+2) or negative (−2) to generate four types of judgments: when the participant responded deceptively to the probe and was wrongly judged to be truthful (positive judgment following probe, PP) or rightly judged to be deceptive (negative judgment following probe, PN) and when the participant responded truthfully to the irrelevant and then was rightly judged to be truthful (positive judgment following irrelevant, IP) or wrongly judged to be deceptive (negative judgment following irrelevant, IN). The judgment design for each type of stimulus is shown in Figure 1A.

Stimulus presentation and behavioral data acquisition were programmed using E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA). The presentations of the stimuli of one trial are illustrated in Figure 1B. At the beginning of each trial, a word was presented on the screen and lasted for 1 s, which was followed by a visual prompt '☆' that lasted for 0.5 s. For each word, the participants were asked to indicate whether they had seen it before by pressing the '1' key using the right thumb to respond 'yes' and '3' using the left thumb to respond 'no'. The assignment of the key to press was counterbalanced among the participants. As instructed by Z.Y.S., the murderers were told to conceal their identities, pretending to know nothing about the crime details, and thus were to deceptively respond 'no' to probes and to truthfully answer 'yes' and 'no', respectively, to targets and irrelevants. However, for innocents, given that they did not commit any crime, they were instructed to simply respond truthfully to all of the stimuli. To reduce the differences in response timing that resulted from the different reading speeds across participants, the participants were asked not to respond until the prompt '☆' appeared (Kozel *et al.*, 2009a). After a varied interval of 0.5/2.5/4.5 s (manipulated as the judging stage), a judgment ('+2'/'−2') appeared for 1.5 s, indicating whether the previous response was judged to be truthful or deceptive,

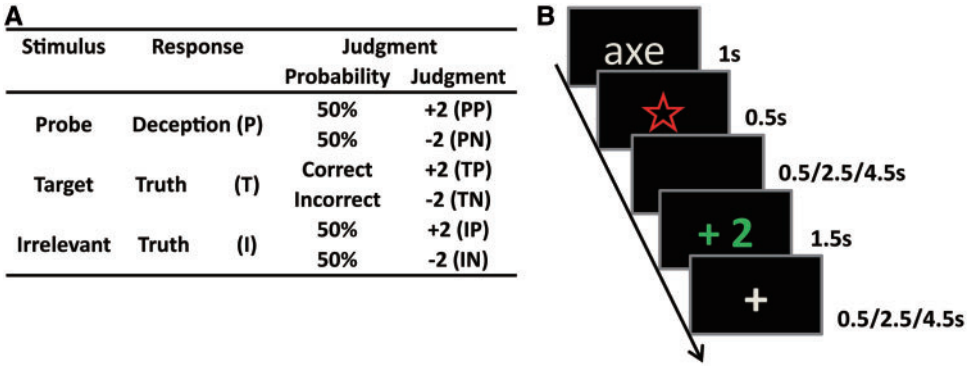

Fig. 1 (**A**) The design of the stimulus presentation and the corresponding judgment given by the computer were equal for both groups. There were three types of stimulus; the judgments for probes and irrelevants were predetermined to be equally likely positive or negative, whereas the judgments to the targets were dependent upon whether the participants responded correctly. (**B**) The presentations of the stimuli of one trial, taking a probe followed by a positive judgment as an example.

and thus two scores were rewarded or punished, followed by a varied inter-trial interval of 0.5/2.5/4.5 s. If the participants pressed an incorrect key, a judgment of '−2' would appear and a warning of 'no response' would be given, which lasted 1.5 s if the participants could not respond within the duration of the prompt '☆' presentation. The participants had sufficient practice with the 'time of the crime' as the stimulus until they achieved 90% accuracy before scanning.

The formal experiment had five sessions, each of which was manipulated to test one type of crime detail (excluding crime time, which had been used in the practice); the order of the sessions was randomized across the participants. Each session included six words: one probe, one target and four irrelevants, with repeated times of 14, 14 and 10, respectively, generating 70 trials in total, which were randomly interspersed. For example, in the session about the weapon, 'dagger' or 'axe' was selected by the murderer in the mock murder phase and was used as the probe, appearing repeatedly 14 times; the target was scissors, repeated 14 times; and the irrelevants included 'hammer', 'kitchen knife', 'gun' and 'iron rod', of which the former two reappeared 10 times and the latter two 11 times.

Post-scan briefing report

After scanning, the participants completed a debriefing report regarding the lie detection process (Supplementary material D).

fMRI data acquisition

Images were acquired using a 3 T Siemens Magnetom Trio Tim B17 MRI scanner equipped with a standard polarized head coil (Siemens Medical Systems, Erlangen, Germany). The T2*-weighted gradient echo planar imaging (EPI) sequences, which were sensitive to the blood oxygenation level-dependent contrast, were used to obtain the functional images of 1360 volumes. Each volume included 32 axial and interleaved acquired slices with 3 mm thickness and 1 mm gap oriented parallel to the AC–PC plane (repetition time = 2000 ms; echo time = 30 ms; field of view = 220 × 220; matrix of 64 × 64; and flip angle = 90°). High-resolution T1-weighted images, which were composed of 176 volumes, were acquired for each participant to be the anatomical reference (repetition time = 1900 ms; echo time = 2.52 ms; slice thickness = 1 mm; field of view = 256 × 256; and voxel size = 1 × 1 × 1 mm).

fMRI data analysis

Data processing and analysis were performed using Statistical Parametric Mapping software (SPM8; Wellcome Department of Cognitive Neurology, London, UK) running with Matlab (Math works, Natick, MA). In the processing stage, slice timing was first conducted to correct the differences in image acquisition time between slices (Ashburner *et al.*, 2011) and realigned to correct for head motion. After being spatially normalized based on the functional EPI template provided by SPM8 and simultaneously resampled to $3 \times 3 \times 3\,mm^3$ resolution, the images were finally smoothed using a Gaussian kernel with a full width at half maximum of 8 mm.

In the first-level analysis, for each participant and session, 11 regressors were modeled using the general linear model in an event-related manner, including three types of stimulus (P, T and I), with onsets corresponding to the stimulus onset for which the correct responses were given by the participants; six types of judgment (PP, PN, TP, TN, IP and IN) following the participants' correct responses, with the onset corresponding to judgment onset; and two special regressors comprising incorrect responses to stimuli (INR) and judgments following these incorrect responses (INJ). To correct for movement-related artifacts, six head-motion parameters from subject-specific realignment were also included in the model. All of the regressors were convolved using the hemodynamic response function, and a high-pass filter set at 128 s was applied to eliminate low-frequency noise. Contrast coefficients were calculated at the first level using *T*-tests, generating statistical parametric maps for each contrast, which were subsequently submitted to group-level random-effect analysis to estimate error variance across individuals.

For the deceptive response stage, we used the contrast of 'P > I' for each participant and also directly compared the brain activations induced by the P condition between the MUD and INC groups to test the brain activity recruited by the deceptive response. For the judgment stage, we used the contrasts of 'PP > IP', 'PN > IN' and their converse contrasts for each participant to test the effect of the deceptive response on brain activity in response to judgment processing. The contrasts of 'PP > PN', 'IP > IN' and their converse contrasts for each participant were calculated to test the valence effect of the judgment. Furthermore, the direct comparisons of the brain activations induced by the PP and PN conditions between the MUD and INC groups were also analyzed. For all of the analyses, the group results were assessed at a threshold of false discovery rate (FDR) < 0.005 and activations involving a contiguous cluster of at least 50 voxels ($K > 50$). For exploratory purposes, we used a more lenient threshold of FDR < 0.05, $K > 50$ to investigate the result of the between-group comparison of the brain activations in response to probes, which showed no significant activations under the threshold of FDR < 0.005, $K > 50$.

To test the accuracy rates of diagnosing whether the participant was truthful or deceptive using the neural signals underlying the deceptive

response and those underlying judgment processing, we conducted region of interest (ROI) analyses using MarsBar (http://marsbar.sourceforge.net/). For the deceptive response stage, four ROIs were functionally defined based on activated regions in the contrast of 'P > I' for the MUD group, including all voxels exceeding the threshold at FDR < 0.005, with no extent threshold (Nose *et al.*, 2009). Subsequently, we conducted discriminant analyses on the bases of the activations of these ROIs in each participant. The independent variable was set to be the parameter estimates for the P condition, and the dependent variable was set to be the groups. We used the stepwise method to select ROIs that contributed to the individual diagnosis with a significant level of enter = 0.05 and stay = 0.10. Finally, to validate the percentage of correct diagnoses, we used the cross-validation method, which leaves one participant who is being diagnosed out of the group analyses for testing. Similarly, for the judgment stage, eight ROIs were functionally defined based on activated regions in the contrast of 'PP > IP' for the MUD group (we considered IP to be the baseline condition of the judgment stage not only because it was the judgment for irrelevant but also because it coincided with the truth; thus, it would not evoke much high or special brain activation), including all voxels exceeding the threshold at FDR < 0.005, with no extent threshold. The same discriminant analysis was conducted, with the expectation that the independent variable was changed to be the parameter estimates of eight ROIs for the PP condition. Using this analysis, ROIs that contributed to the diagnosis were identified, based on which the rates of accurate diagnosis were computed using the cross-validation method. In these analyses, when a participant of the MUD group was diagnosed, the ROIs were redefined based on the data of the participants excluding that one ('leave-one-out' method) (Nose *et al.*, 2009).

RESULTS

Behavioral results

The mean reaction times (RTs) and mean accuracy rates of behavioral response are shown in Table 1. The 2 (group: MUD *vs* INC) × 2 (stimulus: P *vs* I) repeated measures ANOVAs were conducted to analyze RTs and accuracy rates. The RT results showed neither a significant main effect of the group [$F(1, 30) = 0.029$, $P = 0.866$] nor a main effect of the stimulus type [$F(1, 30) = 0.112$, $P = 0.740$], but the interaction effect between these two factors was significant [$F(1, 30) = 6.376$, $P < 0.05$]. Simple effect analyses revealed that in the MUD group, the RT for probes was marginally significantly larger than it was for irrelevants ($P = 0.052$); however, in the INC group, no difference in the RT was found between the probes and the irrelevants ($P = 0.132$). No group difference in the RT was found within the probes or irrelevants ($P = 0.504$ and 0.736, respectively). The behavioral accuracy results showed neither significant main effects of group [$F(1, 30) = 2.469$, $P = 0.127$] nor stimulus type [$F(1, 30) = 0.141$, $P = 0.710$] nor the interaction effect between these two factors [$F(1, 30) = 0.006$, $P = 0.941$].

fMRI results

The regions activated during deceptive response and judgment processing are summarized in Tables 2 and 3.

First, we report the results of the deceptive response stage. In the MUD group, the contrast of 'P > I' activated the bilateral IPLs (BA40), the right VLPF (BA13/45/47) and the right striatum (STR) (Figure 2A). Neither the contrast of 'P > I' for the INC group nor the contrast of 'I > P' for both groups showed significantly activated brain regions. Under a more lenient threshold, we observed that the direct comparison of the brain activation induced by the probes (P) between the MUD and INC groups yielded brain activation patterns that were similar to those induced by the contrast of 'P > I' in the MUD group, which showed stronger activities in the MUD group than in the INC group (Figure 2B), including for the right VLPF extending into the right STR and the right IPL. However, no brain regions showed stronger activity in the INC group than in the MUD group.

Table 1 Mean ± s.d. of RTs and accuracy rates of behavioral responses for probes and irrelevants in the MUD and INC groups

Group	Behavioral measures	Probe	Irrelevant
MUD group	RT	283.011 ± 28.764	276.474 ± 28.103
	Accuracy rate	0.937 ± 0.061	0.941 ± 0.037
INC group	RT	275.344 ± 35.018	280.353 ± 35.839
	Accurate rate	0.910 ± 0.061	0.913 ± 0.057

Second, we report the results of the judgment stage. In the MUD group, the contrast of 'PP > IP' activated a broad brain region (Figure 2C), including the right middle frontal gyrus (MFG), the bilateral IPLs (BA40), the left cerebellum (CER), the bilateral orbitofrontal cortices (OFC) and the bilateral middle temporal gyri (MTG). However, the contrast of 'PP > IP' for the INC group showed no significantly activated brain region. Furthermore, the contrasts of 'PN > IN', 'PP > PN', 'IP > IN' and their converse contrasts for both groups showed no suprathreshold brain activations.

A direct comparison of the brain activation induced by PP between the two groups revealed similar brain activation patterns to those induced by the contrast of 'PP > IP' in the MUD group, which showed stronger activities in the MUD group than in the INC group (Figure 2D), including the bilateral medial superior frontal gyri (mSFG), connected with the left superior frontal gyrus (SFG), the bilateral IPLs (BA40), the bilateral MTG, the bilateral MFGs, the bilateral OFC and the right inferior temporal gyrus (ITG). However, no brain region showed stronger activity in the INC group than in the MUD group for the PP condition. Furthermore, for brain activation induced by PN, neither the contrast of 'MUD > INC' nor the contrast of 'INC > MUD' showed significantly activated regions.

Finally, we reported the results of discriminant analyses, examining which brain regions associated with deceptive response and that associated with judgment processing contributed to diagnosing whether the participant was the murderer or the innocent. In the deceptive response stage, activation in the right VLPF contributed to the individual diagnosis (canonical coefficient = 0.756, eigenvalue = 1.331, Wilks' Lambda = 0.429, $P < 0.001$). Based on the activation in this area for probes (P), we used the cross-validation method to classify each participant into each of the two groups, determining that the accuracy rates of diagnosis were 81.25% for the MUD group and 81.25% for the INC group (total rate, 81.25%). Conversely, in the judgment stage, activation in the right IPL contributed to the individual diagnosis (canonical coefficient = 0.769, eigenvalue = 1.450, Wilks' Lambda = 0.408, $P < 0.001$). Based on the activation in this area for PP, we used cross-validation analyses to classify each participant into each of the two groups, determining that the accuracy rates of diagnosis were 93.75% for the MUD group and 87.50% for the INC group (total accuracy rate, 90.63%). The results of the discriminant analyses are shown in Figure 3 and summarized in Table 4.

DISCUSSION

In this study, after a 'mock murder' situation, we simultaneously examined the brain activation patterns involved in both making a deceptive response and receiving feedback about the effectiveness of the deception using a modified GKT paradigm, in which a judgment

Table 2 Brain regions showed significant activity during the deceptive response stage by whole brain analyses. FDR < 0.005, $K > 50$

	Regions	Side	BA	*t*	Size	MNI		
						x	*y*	*z*
MUD group								
P > I	IPL	L	40	9.47	86	−60	−45	36
	VLPF[a]	R	13/45/47	8.67	275	45	21	3
	IPL	R	40	8.27	81	57	−45	33
	STR	R	—	7.22	75	15	6	0
I > P	No activation							
INC group								
P > I and I > P	No activation							
Between groups[b]								
P: MUD > INC	VLPF	R	13/45/47	6.09	767	33	27	−6
	STR	R	—	5.68	Contiguous with above	15	9	3
	MFG	R	6	5.34	52	42	0	54
	pMFC	R	6	5.13	194	6	21	60
	INS	L	13	4.87	151	−60	15	3
	STR	L	—	4.83	66	−12	6	3
	IPL	R	40	4.59	117	57	−42	33
P: INC > MUD	No activation							

Coordinates refer to the local peak within each cluster.; pMFC, posterior medial frontal cortex; INS, insula; L, left hemisphere; R, right hemisphere.
[a]Area that contributed the most to the diagnosis.
[b]FDR < 0.05, $K > 50$ for exploratory purposes.

Table 3 Brain regions showed significant activity during the judgment stage by whole brain analyses. FDR < 0.005, $K > 50$

	Regions	BA	*t*	Size	MNI		
					x	*y*	*z*
MUD group							
PP > IP	R MFG	8/9	10.47	343	51	30	39
	L IPL	39/40	8.68	667	−57	−51	33
	L CER	—	8.57	308	−27	−81	−27
	L OFC	10/47	8.06	440	−48	39	−15
	R MTG	21	7.95	413	66	−27	−3
	R OFC	10/47	7.87	1071	36	57	−9
	R IPL[a]	39/40	7.77	601	39	−63	57
	L MTG	21	6.53	192	−63	−48	−9
IP > PP	No activation						
PN > IN and PN > IN	No activation						
PP > PN and PN > PP	No activation						
IP > IN and IN > IP	No activation						
INC group							
PP > IP and PP > IP	No activation						
PN > IN and IN > PN	No activation						
PP > PN and PN > PP	No activation						
IP > IN and IN > IP	No activation						
Between groups							
PP: MUD > INC	R mSFG	8/9/10	6.95	1133	9	48	42
	R SFG		6.75	Contiguous with above	21	57	15
	L mSFG		6.45		−12	54	36
	R IPL	40	6.7	526	54	−54	33
	R MTG	21	6.69	475	63	−21	−9
	L IPL	40	6.3	464	−48	−54	27
	L MFG	8	6.27	197	−33	21	42
	R MFG	6/8/9	5.84	247	48	33	39
	R OFC	11/47	5.84	191	39	51	−9
	R ITG	21	5.79	58	48	6	−42
	L OFC	47	5.01	93	−48	36	−9
	L MTG	21	4.69	146	−48	−30	−6
PP: INC > MUD	No activation						
PN: MUD > INC	No activation						
PN: INC > MUD	No activation						

Coordinates refer to the local peak within each cluster. L, left hemisphere; R, right hemisphere.
[a]Area that contributed the most to the diagnosis.

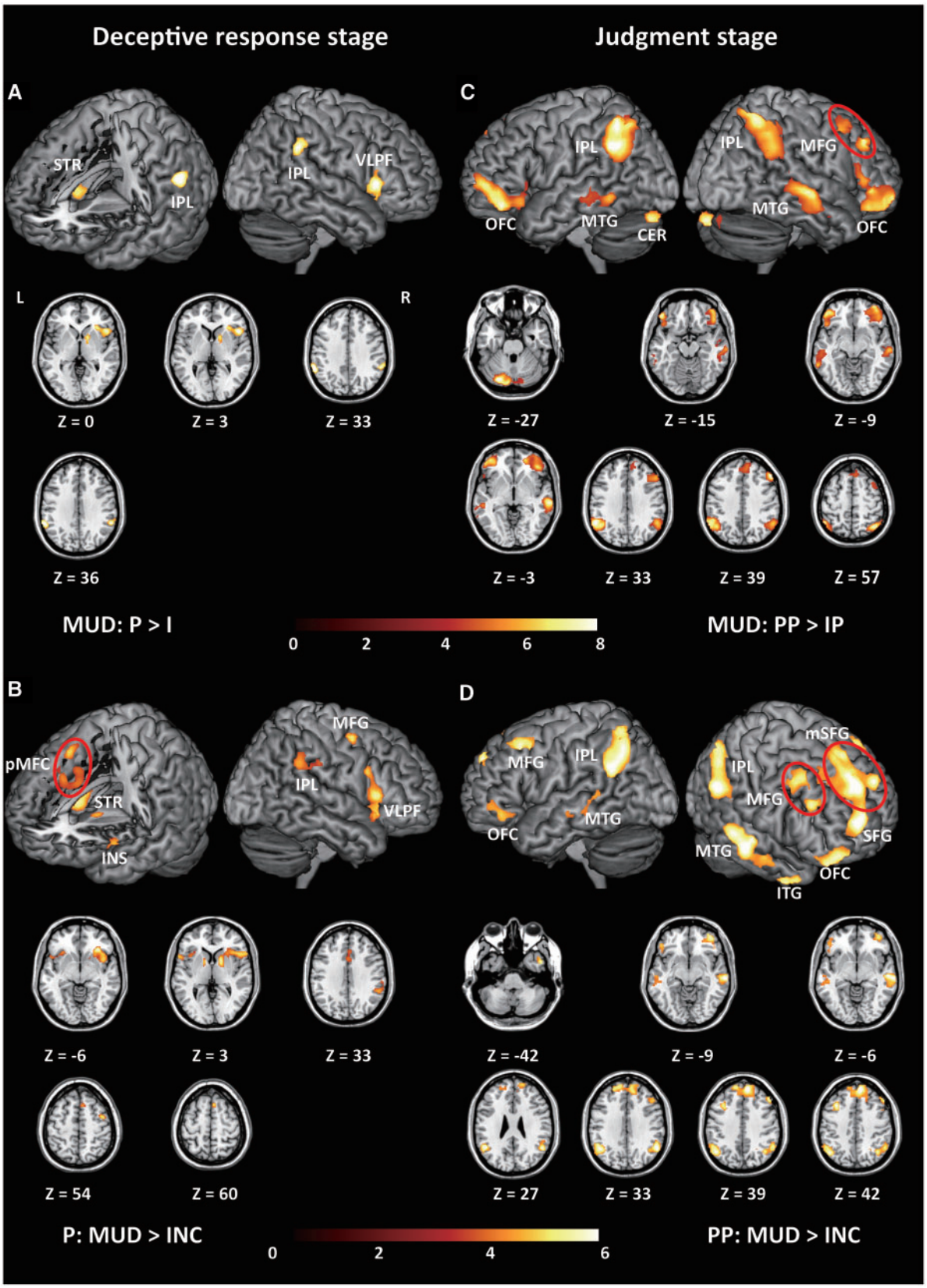

Fig. 2 All of the significant clusters of contrast analyses. When not specified otherwise, data are thresholded at FDR < 0.005, $K > 50$. (**A**) Brain regions showed greater activity for the probes than for the irrelevants in the MUD group. (**B**) Brain regions responding to the probes showed stronger activities in the MUD group than in the INC group. Data are thresholded at FDR < 0.05, $K > 50$ for exploratory purpose. (**C**) Brain regions showed stronger activities for the positive judgment following probes than irrelevants in the MUD group. (**D**) Brain regions responding to positive judgment following probes showed stronger activities in the MUD group than in the INC group. Within each sub-figure, all of the significant clusters were shown on the surface-rendered brain; additionally, they were also shown in axial (*Z*) views at the peak effect coordinates.

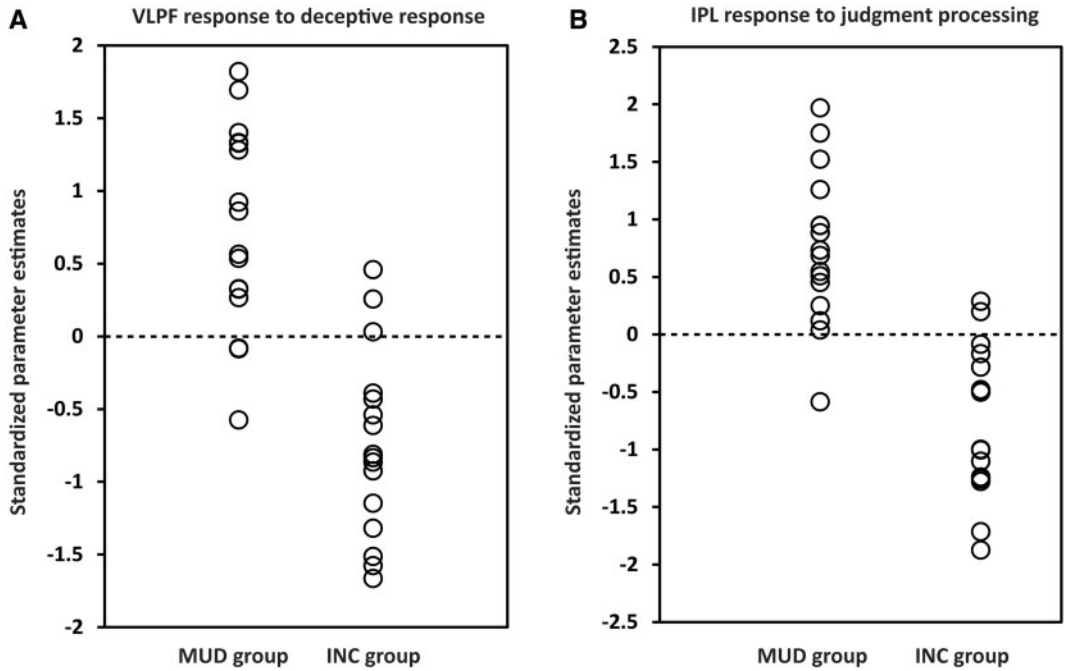

Fig. 3 Individual diagnosis results. (**A**) Results of discriminant analysis based on the activities of the right VLPF in the deceptive response stage. The dots indicate standardized parameter estimates of each participant responding to the probes in the right VLPF. (**B**) Results of discriminant analysis based on the activities of the right IPL in the judgment stage. The dots indicate standardized parameter estimates of each participant responding to the positive judgments following probes in the right IPL. In both (A) and (B), the dotted line represents the value of the threshold for the classification (canonical coefficient, 0.756; eigenvalue, 1.331; Wilks' Lambda, 0.429; $P < 0.001$ for the deceptive response stage, and canonical coefficient, 0.769; eigenvalue, 1.450; Wilks' Lambda, 0.408; $P < 0.001$ for the judgment processing stage).

Table 4 Results of discriminant analyses in the deceptive response stage and the judgment processing stage

	Deceptive response	Judgment processing
Contributing area	Right VLPF	Right IPL
Accuracy rate for MUD	81.25%	93.75%
Accuracy rate for INC	81.25%	87.5%
Total accuracy rate	81.25%	90.63%

indicating whether the participant was judged to be truthful or deceptive was given by the lie detector after receiving the participant's response to each item. The bilateral IPLs (BA40), the right VLPF (BA13/45/47) and the right STR were more active for deceptive responses than for truthful responses. Conversely, for the judgment stage, the right MFG, the bilateral IPLs (BA40), the left CER, the bilateral OFC (BA47) and the bilateral MTG (BA21) were more active for positive judgments following deceptive responses (PP) than for positive judgments following truthful responses (IP), which is considered to be the baseline condition in the judgment stage. Furthermore, we conducted discriminant analyses to examine whether the activation patterns associated with the two processes reliably diagnosed whether the participant was a murderer (deceptive) or an innocent (truthful). We found that for the deceptive response stage, the right VLPF contributed the most to diagnosing the participants' real identities. Based on the activity of this area, 81.25% of the 'murderers' and 81.25% of the 'innocents' were correctly diagnosed (total accuracy rates, 81.25%). For the judgment stage, the right IPL (BA40) contributed the most to the individual diagnosis. Based on the activity of this area, we correctly diagnosed 93.75% of the 'murderers' and 87.5% of the 'innocents' (total accuracy rates, 90.6%). Therefore, the right VLPF may be critical in producing deceptive response, whereas the right IPL may be critical in processing feedback about the effectiveness of the deception.

Neural response to deceptive response

Consistent with previous studies, we found that deceptive responses recruited a set of brain regions that include the bilateral IPLs, the right VLPF and the right STR, which had been previously reported to be involved in deception (Spence *et al.*, 2004; Kozel *et al.*, 2005; Langleben *et al.*, 2005; Fullam *et al.*, 2009; Ganis *et al.*, 2011). As many of these brain regions are involved in much of the executive function paradigm (Spence *et al.*, 2004; Abe, 2011), our findings support the notion that deception demands additional cognitive control processes to suppress a pre-potent truthful response and to produce a deceptive response; thus, deception engages the executive system to a larger extent than does telling the truth, which merely comprises a response in the form of a baseline (Spence *et al.*, 2001, 2004).

Among these regions, the right VLPF was the most critical area for distinguishing deception from the truth. Consistently, the VLPF has been repeatedly reported to be activated during deception (Spence *et al.*, 2001, 2004; Lee *et al.*, 2002; Kozel *et al.*, 2004a,b, 2005; Davatzikos *et al.*, 2005; Luan Phan *et al.*, 2005; Nunez *et al.*, 2005; Gamer *et al.*, 2007; Ganis *et al.*, 2011) and to be the most marked among deception-related brain activation pattern areas (Spence *et al.*, 2001, 2004; Spence and Kaylor-Hughes, 2008; Ganis *et al.*, 2011; Kaylor-Hughes *et al.*, 2011). Considering that the VLPF is characteristically implicated in cognitive control (Ridderinkhof *et al.*, 2004; Blasi *et al.*, 2006; Lie *et al.*, 2006; Dove *et al.*, 2008), the activity in this area observed in this study may support the notion that cognitive control plays a critical role in producing deception and further demonstrates that this control process may be, for the most part, conducted by the right VLPF.

Neural response to judgment processing

The positive judgments following deceptive responses showed obviously stronger activity in a set of brain regions, including the right MFG, the bilateral IPLs, the left CER, the bilateral OFC and the bilateral MTG, compared with the positive judgments following truthful responses. This result is interesting because the two types of positive judgments were the same in their forms (+2), and, more importantly, they would result in a similar consequence (i.e. obtaining two scores). Therefore, it is conceivable that the difference in brain activation between the two types of judgments is attributed to their different subjective levels of significance. Specifically, the positive judgment following deception may be attached to a particularly important implication by the deceiver because it indicated that a deceptive response was wrongly judged to be truthful and therefore comprised more complex information, such as error, conflict and secret joy, than did the positive judgment following truth, which was simply a matter of course. The exact cognitive functions performed by these regions in judgment processing will be investigated in another paper from our research group using a different analytical approach; in this study, however, we focused on the region that is critical for determining whether the participant was truthful or deceptive, and we found that the right IPL contributed the most to correct diagnoses.

The IPL was the most reliable region recruited by deceptive response in the GKT paradigm (Spence, 2008); furthermore, its activation is reportedly one of the most informative neural signals for distinguishing deception from truth (Davatzikos *et al.*, 2005; Langleben *et al.*, 2005). In this study, we demonstrated that this area also plays a critical role in processing judgments about deception. The IPL has been consistently activated in the oddball paradigm, i.e. detecting rare targets from a series of frequent nontargets (Mccarthy *et al.*, 1997; Menon *et al.*, 1997; Linden *et al.*, 1999; Yoshiura *et al.*, 1999; Stevens *et al.*, 2000; Ardekani *et al.*, 2002; Kiehl *et al.*, 2003), even when the participants simply observed infrequent changes in the stimulus without the need to make a response (Downar *et al.*, 2000, 2002). Those studies suggested that the IPL engages in detecting salient stimuli (Downar *et al.*, 2001; Seghier, 2013). In this regard, our findings might reflect that although positive judgments following probes were not infrequent in the entire design, as they had a 50% probability, they had special meaning for the deceivers, which indicated that they had successfully deceived the lie detector. Therefore, these judgments become subjectively salient and might automatically trigger the attentional processes of the deceivers (Cabeza *et al.*, 2008; Petersen and Posner, 2012).

Individual diagnosis based on two-stage brain activation patterns

In this study, we examined the accuracy of the neural signals underlying deceptive response and judgment processing in detecting lies. Among the brain regions associated with deceptive responses, the right VLPF was the area contributing most to individual diagnosis, with 81.25% sensitivity and 81.25% specificity (total accuracy rate, 81.25%). Compared with previous studies that performed individual diagnosis based on brain activation related to deceptive response (Davatzikos *et al.*, 2005; Kozel *et al.*, 2005, 2009a; Langleben *et al.*, 2005), the accuracy rate observed in our study seemed to be approximately average. Conversely, among the brain regions associated with judgment processing, the right IPL contributed, for the most part, to individual diagnosis, with 93.75% sensitivity and 87.5% specificity (total accuracy rate, 90.6%). Compared to the accuracy rates obtained based on the neural signals associated with producing a deceptive response, the accuracy obtained based on the neural signals associated with processing the judgment of the interlocutor appeared to be higher. Therefore, our results suggested that the neural responses involved in processing the effectiveness of one's deception are not only a valid marker of deception but may be more accurate than relying on the neural activity that is associated with producing the deceptive response itself.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Abe, N. (2011). How the brain shapes deception. *Neuroscientist*, *17*, 560–74.

Ardekani, B.A., Choi, S.J., Hossein-Zadeh, G.A., et al. (2002). Functional magnetic resonance imaging of brain activity in the visual oddball task. *Cognitive Brain Research*, *14*, 347–56.

Ashburner, J., Barnes, G., Chen, C., et al. (2011). Functional imaging laboratory: wellcome trust centre for neuroimaging. *SPM8 Manual*. London, UK: 2011.

Blasi, G., Goldberg, T.E., Weickert, T., et al. (2006). Brain regions underlying response inhibition and interference monitoring and suppression. *European Journal of Neuroscience*, *23*, 1658–64.

Cabeza, R., Ciaramelli, E., Olson, I.R., Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*, 613–25.

Davatzikos, C., Ruparel, K., Fan, Y., et al. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, *28*, 663–8.

Dove, A., Manly, T., Epstein, R., Owen, A.M. (2008). The engagement of mid−ventrolateral prefrontal cortex and posterior brain regions in intentional cognitive activity. *Human Brain Mapping*, *29*, 107–19.

Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, *3*, 277–83.

Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *Neuroimage*, *14*, 1256–67.

Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal of Neurophysiology*, *87*, 615–20.

Fullam, R.S., McKie, S., Dolan, M.C. (2009). Psychopathic traits and deception: functional magnetic resonance imaging study. *British Journal of Psychiatry*, *194*, 229–35.

Gamer, M., Bauermann, T., Stoeter, P., Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Human Brain Mapping*, *28*, 1287–301.

Ganis, G., Rosenfeld, J.P., Meixner, J., Kievit, R.A., Schendan, H.E. (2011). Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage*, *55*, 312–9.

Kaylor-Hughes, C.J., Lankappa, S.T., Fung, R., Hope-Urwin, A.E., Wilkinson, I.D., Spence, S.A. (2011). The functional anatomical distinction between truth telling and deception is preserved among people with schizophrenia. *Criminal Behaviour and Mental Health*, *21*, 8–20.

Kiehl, K.A., Laurens, K.R., Duty, T.L., Forster, B.B., Liddle, P.F. (2003). Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology*, *38*, 133–42.

Kozel, F.A., Johnson, K.A., Grenesko, E.L., et al. (2009a). Functional MRI detection of deception after committing a mock sabotage crime. *Journal of Forensic Sciences*, *54*, 220–31.

Kozel, F.A., Johnson, K.A., Mu, Q., Grenesko, E.L., Laken, S.J., George, M.S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, *58*, 605–13.

Kozel, F.A., Laken, S.J., Johnson, K.A., et al. (2009b). Replication of functional MRI detection of deception. *Open Forensic Science Journal*, *2*, 6–11.

Kozel, F.A., Padgett, T.M., George, M.S. (2004a). A replication study of the neural correlates of deception. *Behavioral Neuroscience*, *118*, 852–6.

Kozel, F.A., Revell, L.J., Lorberbaum, J.P., et al. (2004b). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *Journal of Neuropsychiatry and Clinical Neurosciences*, *16*, 295–305.

Langleben, D.D. (2008). Detection of deception with fMRI: are we there yet? *Legal and Criminological Psychology*, *13*, 1–9.

Langleben, D.D., Loughead, J.W., Bilker, W.B., et al. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *26*, 262–72.

Langleben, D.D., Schroeder, L., Maldjian, J.A., et al. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage*, *15*, 727–32.

Lee, T.M.C., Liu, H.-L., Tan, L.-H., et al. (2002). Lie detection by functional magnetic resonance imaging. *Human Brain Mapping*, *15*, 157–64.

Lie, C.H., Specht, K., Marshall, J.C., Fink, G.R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *Neuroimage*, *30*, 1038.

Linden, D.E.J., Prvulovic, D., Formisano, E., et al. (1999). The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cerebral Cortex*, *9*, 815–23.

Luan Phan, K., Magalhaes, A., Ziemlewicz, T.J., Fitzgerald, D.A., Green, C., Smith, W. (2005). Neural correlates of telling lies: a functional magnetic resonance imaging study at 4 Tesla1. *Academic Radiology*, *12*, 164–72.

Lynoe, N., Sandlund, M., Dahlqvist, G., Jacobsson, L. (1991). Informed consent: study of quality of information given to participants in a clinical trial. *British Medical Journal*, *303*, 610–613.

Mccarthy, G., Luby, M., Gore, J., Goldman-Rakic, P. (1997). Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *Journal of Neurophysiology*, *77*, 1630–4.

Menon, V., Ford, J.M., Lim, K.O., Glover, G.H., Pfefferbaum, A. (1997). Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *Neuroreport*, *8*, 3029–37.

Nose, I., Murai, J., Taira, M. (2009). Disclosing concealed information on the basis of cortical activations. *Neuroimage*, *44*, 1380–6.

Nunez, J.M., Casey, B., Egner, T., Hare, T., Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage*, *25*, 267–77.

Petersen, S.E., Posner, M.I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73.

Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.

Ridderinkhof, K.R., van den Wildenberg, W.P.M., Segalowitz, S.J., Carter, C.S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*, 129–40.

Seghier, M.L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist*, *19*, 43–61.

Sip, K.E., Roepstorff, A., McGregor, W., Frith, C.D. (2008). Detecting deception: the scope and limits. *Trends in Cognitive Sciences*, *12*, 48–53.

Sip, K.E., Skewes, J.C., Marchant, J.L., McGregor, W.B., Roepstorff, A., Frith, C.D. (2012). What if I get busted? Deception, choice, and decision-making in social interaction. *Frontiers in Neuroscience*, *6*, 58.

Spence, S.A. (2008). Playing Devil's advocate: the case against fMRI lie detection. *Legal and Criminological Psychology*, *13*, 11–25.

Spence, S.A., Farrow, T.F.D., Herford, A.E., Wilkinson, I.D., Zheng, Y., Woodruff, P.W.R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, *12*, 2849–53.

Spence, S.A., Hunter, M.D., Farrow, T., et al. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*, 1755–62.

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.

Stevens, A.A., Skudlarski, P., Gatenby, J.C., Gore, J.C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magnetic Resonance Imaging*, *18*, 495–502.

Wolpe, P.R., Foster, K.R., Langleben, D.D. (2005). Emerging neurotechnologies for lie-detection: promises and perils. *American Journal of Bioethics*, *5*, 39–49.

Yoshiura, T., Zhong, J., Shibata, D.K., Kwok, W.E., Shrier, D.A., Numaguchi, Y. (1999). Functional MRI study of auditory and visual oddball tasks. *Neuroreport*, *10*, 1683–88.