NeuroImage xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging

Giorgio Ganis [a,b,c,*], J. Peter Rosenfeld [d], John Meixner [d], Rogier A. Kievit [e], Haline E. Schendan [b,c]

[a] Department of Radiology, Harvard Medical School, Boston, MA 02115, USA
[b] Massachusetts General Hospital, Martinos Center, Charlestown, MA 02129, USA
[c] School of Psychology, University of Plymouth, Plymouth, Devon, PL48AA, UK
[d] Department of Psychology, Northwestern University, Evanston, IL 60208-2710, USA
[e] Department of Psychology, University of Amsterdam, Amsterdam, 1018WB, Netherlands

ARTICLE INFO

Article history:
Received 7 September 2010
Revised 27 October 2010
Accepted 5 November 2010
Available online xxxx

ABSTRACT

Functional magnetic resonance imaging (fMRI) studies have documented differences between deceptive and honest responses. Capitalizing on this research, companies marketing fMRI-based lie detection services have been founded, generating methodological and ethical concerns in scientific and legal communities. Critically, no fMRI study has examined directly the effect of countermeasures, methods used by prevaricators to defeat deception detection procedures. An fMRI study was conducted to fill this research gap using a concealed information paradigm in which participants were trained to use countermeasures. Robust group fMRI differences between deceptive and honest responses were found without, but not with countermeasures. Furthermore, in single participants, deception detection accuracy was 100% without countermeasures, using activation in ventrolateral and medial prefrontal cortices, but fell to 33% with countermeasures. These findings show that fMRI-based deception detection measures can be vulnerable to countermeasures, calling for caution before applying these methods to real-world situations.

Introduction

Deception is a pervasive behavior that can serve useful social purposes (DePaulo et al., 1996) but can also have enormous negative consequences, which is why societies have long sought reliable methods for determining when people lie (Vrij, 2008). Methods have included observing behavioral and peripheral physiology (Vrij, 2008). To improve upon these methods (National Research Council, 2003), researchers recently began monitoring brain activity with event-related brain potentials (ERPs) and, lately, with functional magnetic resonance imaging (fMRI). FMRI laboratory studies have shown that deceptive and honest responses can be differentiated in group data and intraindividual analyses have revealed deception detection accuracies around 90% (e.g., Abe et al., 2008; Bhatt et al., 2008; Davatzikos et al., 2005; Gamer et al., 2007, in press; Ganis et al., 2003, 2009; Kozel et al., 2004, 2005; Langleben et al., 2002, 2005; Lee et al., 2005, 2009; Mohamed et al., 2006; Monteleone et al., 2008; Nose et al., 2009; Nunez et al., 2005; Spence et al., 2001, 2008). Capitalizing on this research, companies have begun marketing fMRI-based "lie detection" services, capturing the imagination of the popular media, but generating methodological and ethical concerns in scientific and legal communities (Greely and Illes, 2007; Nature Neuroscience Editorial, 2008). One such concern is that the accuracy of current fMRI-based methods for real world applications may be overestimated by the public because neuroimaging data are typically perceived by non-experts as being more compelling than other types of data (Weisberg et al., 2008). A virtually unexplored aspect of this concern is whether *countermeasures*, methods prevaricators employ to confuse deception detection procedures, could defeat fMRI deception tests. This issue is critical because countermeasures are known to degrade the accuracy of deception detection using peripheral physiological and ERP measurements (Honts et al., 1996; Rosenfeld et al., 2004).

To address this issue, we conducted an fMRI study using a modified concealed information test (CIT, also referred to as "guilty knowledge test") in which participants were trained to use a covert countermeasure while lying about knowing their birth date. Methodologically, CIT paradigms are the gold standard in laboratory research to determine if a person is lying about possessing knowledge of an item of interest (or "probe") (Ben-Shakhar and Elaad, 2003). CIT paradigms rely on the finding that a salient stimulus, such as an infrequent and meaningful item presented within a series of nonsalient items, produces an orienting response that directs attention to potentially important changes in the environment (Lykken, 1974). This oddball response is greater if it is the only one associated with deception. By using appropriate nonsalient comparison items (or "irrelevants"), this

* Corresponding author. Martinos Center, Building 149, Massachusetts General Hospital, Charlestown, MA 02129, USA. Fax: +1 617 812 0524.
E-mail address: ganis@nmr.mgh.harvard.edu (G. Ganis).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.11.025

response can be used to infer that a person possesses knowledge about a probe but deceptively reports no such knowledge. Individuals with no knowledge about the probe, and who truthfully claim so, will show a much smaller response.

The CIT protocol employed here has been called the "3-stimulus" protocol in the ERP literature (Winograd and Rosenfeld, in press) since it contains on any trial either a probe, irrelevant or an attention holding "target" (an irrelevant item to which participants are assigned a unique response, as articulated in more detail in the Design and procedure section). This protocol has been used both in fMRI (e.g., Nose et al., 2009; Gamer et al., 2007) and ERP (e.g., Rosenfeld et al., 2004) work. FMRI studies using variants of this protocol have reported stronger activation to probes than irrelevants in regions including the lateral and medial prefrontal cortex (e.g., Langleben et al., 2002; Nose et al., 2009; Phan et al., 2005; Gamer et al., 2007). Such activations have been attributed to memory-related and executive control processes (Christ et al., 2009) that are likely to be engaged more strongly by probes (especially when they require a deceptive response) than by irrelevants. Critically, ERP studies using the 3-stimulus protocol and focusing on the P300 potential have shown that this protocol is vulnerable to countermeasures in which participants covertly assign meaning to the nonsalient comparison stimuli in order to reduce the relative salience of the probe (Rosenfeld et al., 2004, also replicated in recent work of ours to be published elsewhere). The key question examined here is whether these same countermeasures can also decrease the accuracy of a 3-stimulus CIT paradigm using fMRI measures of brain activation. This is an important question because fMRI responses to probe items may not index the same brain activity underlying the P300 to these same items, and so it is possible that an fMRI-based CIT protocol using the 3-stimulus protocol might not be susceptible to the countermeasures used in the ERP studies.

Note that Rosenfeld et al. (2008) recently developed a new ERP-based "Complex Trial" protocol that is more resistant to countermeasures and that might have been used here for fMRI. However, this protocol cannot be easily adapted to fMRI, given the slow hemodynamic nature of the signals it measures, because it requires the presentation of stimuli in rapid succession during each trial. Therefore, we chose to start with the simpler 3-stimulus protocol here so as to answer the basic empirical question of its vulnerability to countermeasures using fMRI signals.

Materials and methods

Subjects

Twenty-six Harvard University undergraduates (14 females; mean age = 20.1 years) participated. All gave written informed consent following protocols approved by the Massachusetts General Hospital and Harvard University Institutional Review Boards. Twelve participants were employed in the main study (main group). Fourteen others were included in a second group (ROI group) to obtain independent ROIs of which two did not complete the study, due to technical problems, and their data were not used. All participants had normal or corrected-to-normal vision, no history of neurological disease, and were right-handed. Participants were paid a base rate of $100 plus a $20 bonus.

Stimuli

In each condition, the stimuli were 6 dates shown in white against a black background (1 by 6° of visual angle) and presented for 500 ms as shown in Fig. 1. The stimuli were followed by a black screen with a fixation dot lasting between 1500 and 9500 ms (2500 ms, on average), according to a pseudo-random sequence (Dale, 1999).

Design and procedure

Tasks were presented using Psyscope X on a MacBook. The three conditions were (i) no knowledge (NK), (ii) concealed knowledge (CK), and (iii) countermeasure (CM). All conditions included three types of items (dates) as follows. Four "irrelevant" dates (66.7% of all stimuli), different in each condition, had no particular meaning for participants, who responded truthfully by pressing the "no" button with their right hand, indicating that they did not know these items. Participants also saw one infrequent "probe" date (16.7% of all stimuli). However, the meaning of this probe date and the response instructions varied by condition. For the no knowledge condition, the probe date was just another irrelevant date without any particular meaning for the participants, who simply pressed the "no" key to this item, indicating truthfully that they did not know it; participants received no information about this irrelevant probe date beforehand. Hence, this control condition simulated the case of participants without concealed knowledge about the probe. In contrast, for the concealed knowledge condition, the probe was the participant's birth date, and participants were instructed to lie about whether they knew this date (i.e., as meaningful because it was their birth date) by pressing the "no" button (Fig. 1). The countermeasure condition was the same as the concealed knowledge condition, except participants were instructed how to perform a countermeasure consisting in associating distinct covert actions to 3 of the 4 irrelevants in the sequence and in performing such covert actions each time they saw the associated irrelevant. The 3 covert actions were: to move imperceptibly (i.e., without any overt movement that could be observed) the left index finger, the middle left finger, and the left toe. No covert action was used for the fourth irrelevant date so that exactly half of all items were associated with covert actions. Participants were asked to perform the countermeasure just before pressing the button to indicate whether they knew the dates, enabling us to confirm they had engaged in the countermeasure by examining the response time (RT) pattern (as in Rosenfeld et al., 2008; Rosenfeld and Labkovsky, 2010). Similarly, these countermeasures were designed to have an implicit motor component so their deployment could be demonstrated by examining the pattern of fMRI activation in the motor cortex. Note that participants pressed the "no" button to the irrelevants and probes in all conditions, so that any differences between these items cannot be attributed to response differences. Finally, in all conditions, participants also saw an infrequent "target" date (16.7% of all stimuli), to which they responded by pressing the "yes" button with their right hand, indicating that they knew this date. This target date was studied by participants before the fMRI session, and it was included to ensure attention was paid in all conditions: without this item, one could perform the task by mindlessly responding "no" to all items.

To emphasize the social nature of the task, participants were told that an observer outside the scanner would monitor their eye movements and facial expressions. After the no knowledge condition requiring no deception, but before the concealed knowledge and countermeasure conditions, participants were told that this observer would try to determine if they were lying. They were also told that they would receive a 50c bonus for each trial in which they successfully lied about their birth date and mislead the external observer. At the end of the study, each participant was given a $20 bonus.

To ensure that no irrelevant date was salient to participants, during the week preceding the study, participants provided a list of dates that had special meaning to them. Irrelevant dates were chosen from months and days different from any of these dates, or from the target date. No irrelevant date coincided with common salient dates (e.g., the 25th of December). Before the study, participants were given a set of 45 dates, including the dates they had indicated were meaningful to them (randomly interspersed with the other dates),

Please cite this article as: Ganis, G., et al., Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.11.025

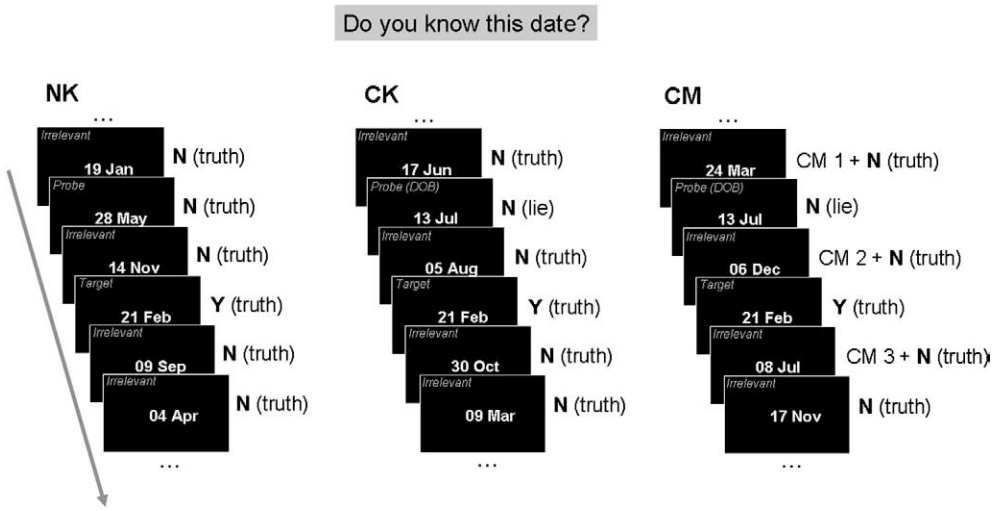

Fig. 1. Concealed information task paradigm. Schematic of stimuli employed in the *no knowledge* (NK), *concealed knowledge* (CK), and countermeasure (CM) conditions. Stimuli included *irrelevant* dates and an infrequent *probe* date. Irrelevant dates were nonsalient dates with no particular meaning to participants. In the no knowledge condition, the probe was an additional irrelevant date; hence, in this condition, participants had no knowledge about the probe date. In the concealed knowledge condition, the probe was the birth date of each participant. There was also a third type of stimulus, an infrequent *target* date studied before the fMRI session, to ensure that participants had to attend the stimuli to perform the task. Participants responded truthfully to all irrelevant and target dates ("no" and "yes", respectively) and deceptively ("no") to the probe date. The countermeasure condition was the same as the concealed knowledge condition, but participants performed 3 distinct countermeasures on 3 of the irrelevant dates, just before indicating whether they knew the dates.

and they were asked to cross out any dates they knew. All dates described as salient earlier were also crossed out by participants during this verification step.

In the main group, participants were tested in the no knowledge, concealed knowledge, and countermeasure conditions. The no knowledge condition was administered before the concealed knowledge condition because after performing this condition (which required lying) participants would have become aware of the purpose of the study and might have not been able to act as individuals with no knowledge about the probe. Furthermore, the countermeasure condition was administered last to ensure people would not use systematic countermeasures during the concealed condition, which they might have done if this condition followed the countermeasure condition. Only the concealed knowledge condition was administered to participants in the ROI group. Five blocks of 36 trials were used for each condition.

Prior to the MRI session, a health history was administered. Participants read instructions on the computer screen and paraphrased them aloud. We corrected any misconceptions at this time. Responses were given using an MRI-compatible button box with two keys. Participants were instructed to respond as quickly as possible without sacrificing accuracy. They were instructed to fixate their gaze on the center of the screen at all times. Ten practice trials were administered before the no knowledge and concealed knowledge conditions. Additional practice (a total of 36 trials) was given just before the countermeasure condition, because someone who intends to use the countermeasures would be likely to practice them at length and we wanted our countermeasures to be as effective as possible.

MRI methods and analyses

A Siemens 3T MAGNETOM TIM Trio whole-body MR scanner with a standard head coil was used. A high-resolution anatomical volume was acquired after the functional scans and the SPGR pulse sequence. Functional scans assessed blood oxygenation changes, using a T2*-sensitive sequence (gradient echo, TR = 2000 ms, TE = 30 ms, FOV = 20 cm, flip angle = 90°, 64 × 64 matrix, voxel size = 3.125 × 3.125 × 4 mm). Each volume in the functional scans was composed of 32, near-axial slices. The stimuli were projected via a magnetically shielded LCD video projector onto a translucent screen placed behind the head of participants. Participants saw the screen via a front-surface mirror on the head coil.

Images were analyzed with AFNI (Cox, 1996) as follows: (a) slice timing correction; (b) motion correction; (c) spatial smoothing with a Gaussian filter (full-width half-maximum = 6 mm); (d) amplitude normalization, by scaling timeseries to a mean of 100 and calculating the percent signal change about this mean; (e) spatial normalization to the MNI305 template; and (f) spatial resampling to a 3 × 3 × 3 mm grid. For the hemodynamic response function, a gamma-variate model was used, with amplitude estimated using multiple linear regression. The multiple regression model included linear, quadratic and cubic trend regressors for each scan to model slow signal drifts. Each item type was modeled by a separate regressor. Incorrect trials were modeled by an additional regressor but not analyzed further. Maps of percent signal change for each participant and condition were obtained using the corresponding regression coefficients.

The primary whole-brain analysis was a paired *t*-test in the main group, comparing the activation between the probe and the mean of the four irrelevants, $p < 0.01$, FDR corrected, (Genovese et al., 2002), in the concealed knowledge condition. The same whole-brain analysis was carried out in the ROI group. Clusters in the ROI group with an overlap of 40 or more voxels with a cluster in the main group were used to define the ROIs for the subsequent analyses. This ROI definition procedure avoided circularity (Kriegeskorte et al., 2009) because the datasets used to define the ROIs and to perform statistics were independent. ANOVAs were carried out on the data from the main group employing these independently defined ROIs, using the factors of condition, item type, and ROI. These ROIs were also used for the single subject classification analyses.

Since participants performed covert actions with their left index and middle fingers upon seeing the first two irrelevants during the countermeasure condition, we expected increased activation in the contralateral (right) motor cortex for these items, compared to the fourth irrelevant (not associated with countermeasures). Thus, we analyzed the activation in the primary motor representation of the left hand fingers using a spherical ROI (radius = 7 mm), centered at published Talairach coordinates (x = 36, y = -22, z = 58)(Alkadhi et al., 2002).

A known issue in the machine learning field is that perfect classification performance can be achieved in the absence of

Please cite this article as: Ganis, G., et al., Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.11.025

generalization because of data overfitting (Hand et al., 2001). Hence, to assess single subject classification performance as well as generalization, we used a jackknife method (Efron, 1982). A binary support vector machine (SVM) with linear kernel (using the Statistical Pattern Recognition Toolbox for Matlab, STPRTool) was trained on data from all possible subsets of 11 participants (each with 22 cases, 11 from the concealed knowledge and 11 from the corresponding no knowledge condition) and tested on the data from the remaining participant (one case from each of the concealed knowledge, no knowledge, and countermeasure conditions). This procedure is similar to the typical way in which analyses would be performed in real-life cases: a model is built on a training group of no knowledge and concealed knowledge cases and tested on a different group of cases (Kozel et al., 2005). The difference here is that some cases within the testing group are countermeasure cases and the critical question is the extent to which they are correctly classified as concealed knowledge cases or incorrectly as no knowledge cases. For each case and condition, the difference was computed between the number of voxels activated by the probe and irrelevants within the independently defined ROIs. These numbers were used as inputs to the classifier. The SVM classifier determined the parameters of the hyperplane that discriminated the no knowledge and concealed knowledge cases with the largest possible margin (i.e., with the largest distance to the nearest training data points from the two conditions). Classification accuracy was measured by the proportion of test cases falling on the correct side of the hyperplane, with perfect accuracy corresponding to the situation of all no knowledge cases falling on one side of the hyperplane and all concealed knowledge cases falling on the other side. We determined the classification accuracy for all individual ROIs and for all possible ROI pairs and triplets and the average margin (across all 12 training subsets tested with the same ROIs) was calculated. Since perfect no knowledge/concealed knowledge classification was achieved with 3 ROIs or fewer, and since we had only a limited number of training examples, combinations of more than 3 ROIs were not tested, to minimize the risk of overfitting. For a similar reason, since perfect performance was achieved with linear SVMs, generalization was not tested with non-linear SVM classifiers.

Results

Behavioral results

To confirm that participants performed the tasks as instructed, first we used one-sample *t*-tests on error data for all item types and conditions. This analysis showed that accuracy was well above the 50% chance level, all $ts(11)>11.5$, all $ps<0.0001$, with the lowest accuracy at 86.5% for targets in the concealed knowledge condition. Next, planned *t*-tests on the RT and error data showed slower responses for probes (M = 748 ms, SE = 39 ms) than irrelevants (M = 688 ms, SE = 38 ms) in the concealed knowledge condition, $t(11)=2.4$, $p<0.05$. Similarly, error rates were higher for probes (M = 3.6%, SE = 1.3%) than irrelevants (M = 0.3%, SE = 0.2%), $t(11)=2.5$, $p<0.005$. In contrast, no RT differences between probes (M = 712 ms, SE = 40 ms) and irrelevants (M = 716 ms, SE = 39 ms) were found in the control no knowledge condition, $t(11)=0.36$, $p>0.1$. Similarly, no error rate differences were found between probes (M = 1.4%, SE = 0.8%) and irrelevants (M = 1.5%, SE = 0.5%), $t(11)=0.1$, $p>0.1$. Thus, processing probes in the concealed knowledge condition, which required deceptive responses, was more costly than processing irrelevants requiring an honest response, as found in other studies (e.g., Seymour et al., 2000). In the countermeasure condition, RTs were slower for irrelevants (M = 1254 ms, SE = 46 ms), than probes (ME = 941 ms, SE = 58 ms), $t(11)=8.9$, $p<0.001$. Error rates did not differ between irrelevants (M = 1.1%, SE = 0.5%) and probes (M = 1.3%, SE = 0.6%), $t(11)=0.2$, $p>0.1$. The slower RTs for irrelevants in this condition confirmed that participants performed the countermeasures before indicating whether they knew the dates, as instructed.

FMRI results

Group analyses

A paired *t*-test comparing brain activation between probes and irrelevants in the concealed knowledge condition (main group) revealed 14 significant activation clusters (Fig. 2, Table 1). Seven of these clusters, the largest ones, overlapped with those found in the independent ROI group (Fig. 3) used to define ROIs for the subsequent analyses. Two ANOVAs were conducted on these seven ROIs using the within-subject factors of item type (irrelevants vs. probes), condition (concealed knowledge and no knowledge, or concealed knowledge and countermeasure), and ROI. The first analysis included the concealed and no knowledge conditions and showed a main effect of condition, $F(1,11)=36.2$, $p<0.001$, $\eta_P^2=0.77$, and of item type, $F(1,11)=134.1$, $p<0.001$, $\eta_P^2=0.92$. As expected, the difference between probes and irrelevants was larger in the concealed than in the no knowledge condition, shown by the interaction between item type and condition, $F(1,11)=68.2$, $p<0.001$, $\eta_P^2=0.86$, and this interaction was modulated by ROI, $F(6,66)=4.9$, $p<0.005$, $\eta_P^2=0.31$. The second analysis included the concealed knowledge and countermeasure conditions. Results showed a main effect of condition, $F(1,11)=10.4$, $p<0.01$, $\eta_P^2=0.49$, and of item type, $F(1,11)=75.6$, $p<0.001$, $\eta_P^2=0.87$. Critically, the difference between probes and irrelevants was larger in the concealed knowledge than in the countermeasure condition, as shown by the interaction between item type and condition, $F(1,11)=176.4$, $p<0.001$, $\eta_P^2=0.94$. The effect of item type also varied by ROI, $F(6,66)=6.1$, $p<0.005$, $\eta_P^2=0.36$. In sum, the difference between probes and irrelevants, the primary index of deception in this paradigm, was largest in the concealed knowledge condition.

Single subject classification

The independently defined ROIs were used also for individual diagnosis. Results showed that two ROIs, each by themselves, could classify cases from the no knowledge and concealed knowledge conditions with 100% accuracy: the right lateral prefrontal (GFi/INS) and the anterior medial prefrontal cortex (GC/GFs/GFd). These rates are somewhat higher than those reported in previous work using activation from similar brain regions (Davatzikos et al., 2005; Kozel et al., 2005; Nose et al., 2009), probably because of the highly salient autobiographical probes used here. Combining these two ROIs increased the robustness of the classification, as indexed by the margin associated with the trained classifier (i.e., distance of the classification hyperplane to the nearest training cases from the no knowledge and concealed knowledge conditions). Finally, the ROI triplet with the largest margin of all possible triplets included the left and right lateral prefrontal and the anterior medial prefrontal regions (Fig. 4a). This ROI triplet was employed to test the effect of the countermeasure on the accuracy of single subject classification. Results showed a lower correct classification rate for countermeasure than concealed knowledge cases (Fig. 4b): classification performance on countermeasure cases was significantly worse than on concealed knowledge cases (4 vs. 12 out of 12), $\chi^2(1)=9.2, p<0.005$. The countermeasure caused most deceptive cases to be classified as honest cases (false negatives).

Potential effect of practice and habituation

The reduced classification rates in the countermeasure condition (always tested last because of the within-subject design) could be due in part to practice effects or to habituation. If this was the case, then activation differences between probes and irrelevants in the 3 ROIs that produced the best classification rates between the concealed knowledge and no knowledge conditions should decrease significantly over time. An ANOVA conducted on the 5 blocks in the

Please cite this article as: Ganis, G., et al., Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.11.025

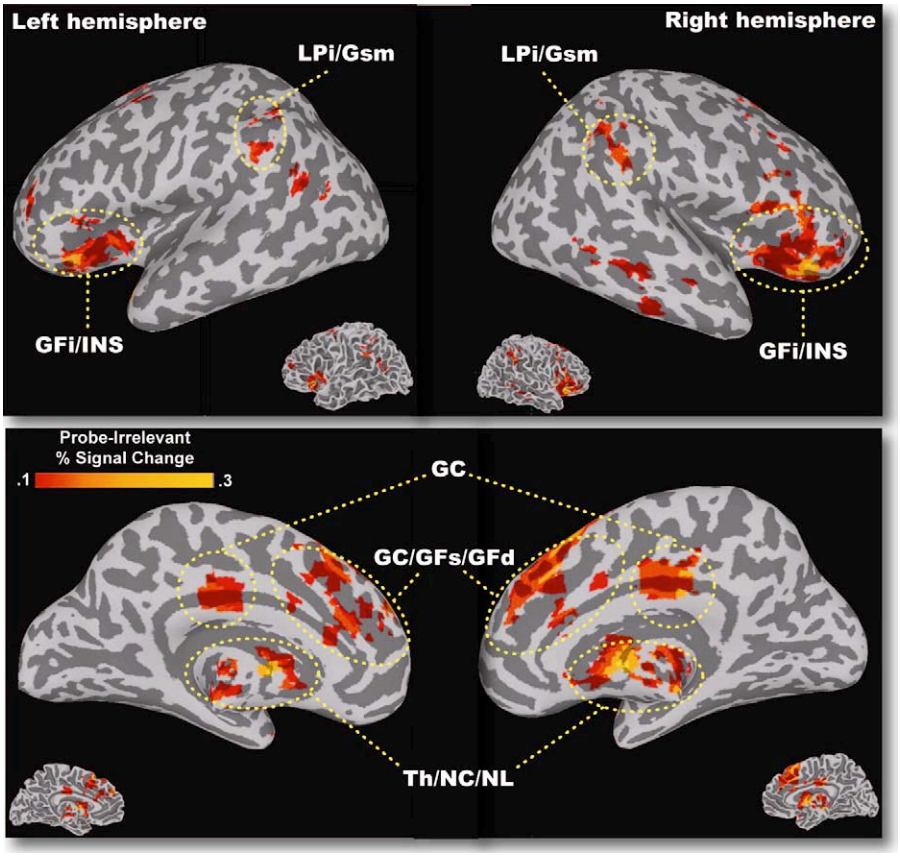

Fig. 2. Differences between probes and the mean of irrelevant items in the main group ($n = 12$) for the concealed knowledge condition ($p < 0.01$, FDR corrected for multiple comparisons), shown on an inflated brain (top: lateral view; bottom: medial view). The color scale depicts percent signal change. The seven activation clusters labeled and indicated by yellow ellipses were found also using the same contrast in the ROI group. Note that the 3 medial regions were combined into single bilateral clusters. Abbreviations for the brain region labels are as in Table 1 (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

concealed knowledge condition using block as factor (collapsing across ROIs) showed this not to be the case, $F(4,44) = 1.03$, $p > 0.1$, $\eta^2_P = 0.094$ (Block 1: M = 0.29, SE = 0.048; Block 2: M = 0.23, SE = 0.053; Block 3: M = 0.25, SE = 0.048; Block 4: M = 0.30, SE = 0.027; Block 5: M = 0.21, SE = 0.037). The same analysis conducted on the 5 blocks in the countermeasure condition found no systematic activation decrease as a function of block, $F(4,44) = 0.50$, $p > 0.1$, $\eta^2_P = 0.047$ (Block 1: M = 0.044, SE = 0.026; Block 2: M = 0.049, SE = 0.027; Block 3: M = 0.068, SE = 0.029; Block 4: M = 0.050, SE = 0.028; Block 5: M = 0.048, SE = 0.030).

Moreover, using a 3-stimulus protocol of the type used here, Rosenfeld et al. (2007) found no changes in the Probe-Irrelevant P300 differences over three repeated blocks with the concealed knowledge participants.

Table 1
Brain Regions Showing Stronger Activation for Probe than Irrelevant Dates in the CK Condition.

Regions within cluster (BA)	Volume	Talairach coordinates (center of mass)			Talairach coordinates (range)					
		x	y	z	Min x	Max x	Min y	Max y	Min z	Max z
GC/GFs/GFd (32/33/6/24/8)	19,980	5	23	45	−12	36	0	51	18	72
GC (23/24)	2808	1	−18	32	−12	9	−27	−6	27	39
GFi/INS (47/45/44)	16,362	43	23	−1	24	63	6	45	−24	24
GFi/INS (47/45)	7884	−37	18	−3	−51	−24	6	30	−24	12
GFm/GFd/GPrC (6)	2538	−23	2	57	−33	−12	−9	12	45	69
GFm/GFi/GPrC (6/9/8)	2025	46	12	39	39	57	6	18	24	51
GFm/GFs (10/9)	1242	−36	48	21	−42	−33	42	57	12	30
LPi/Gsm (40)	3483	53	−43	40	42	63	−54	−33	27	57
LPi/Gsm (40)	2106	−59	−42	34	−63	−51	−51	−33	21	45
LPi/Gsm (40)	567	−44	−46	47	−51	−39	−51	−45	39	54
GTm/GTi (21)	1647	53	−26	−11	45	63	−36	−15	−15	−6
GTs/GTm (39/22)	1512	−51	−58	17	−60	−42	−66	−51	12	24
GTm/GTi (21/37)	837	62	−47	−2	57	66	−54	−42	−6	3
Th/NC/NL	12,555	2	−4	5	−18	21	−30	15	−12	18

Note. In bold are regions that were also identified in the ROI group (Fig. 3). Abbreviations: BA, Brodmann's area; GC, cingulate gyrus; GFs, superior frontal gyrus; GFd, medial frontal gyrus; GFi, inferior frontal gyrus; INS, insula; GPrC, precentral gyrus; GFm, middle frontal gyrus; LPi, inferior parietal lobule; Gsm, supramarginal gyrus; GTm, middle temporal gyrus; GTi, inferior temporal gyrus; Th, thalamus; NC, caudate nucleus; and NL, lenticular nucleus. Coordinates: x, left/right; y posterior/anterior; z, inferior/superior. Volume is in mm^3.

Please cite this article as: Ganis, G., et al., Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.11.025

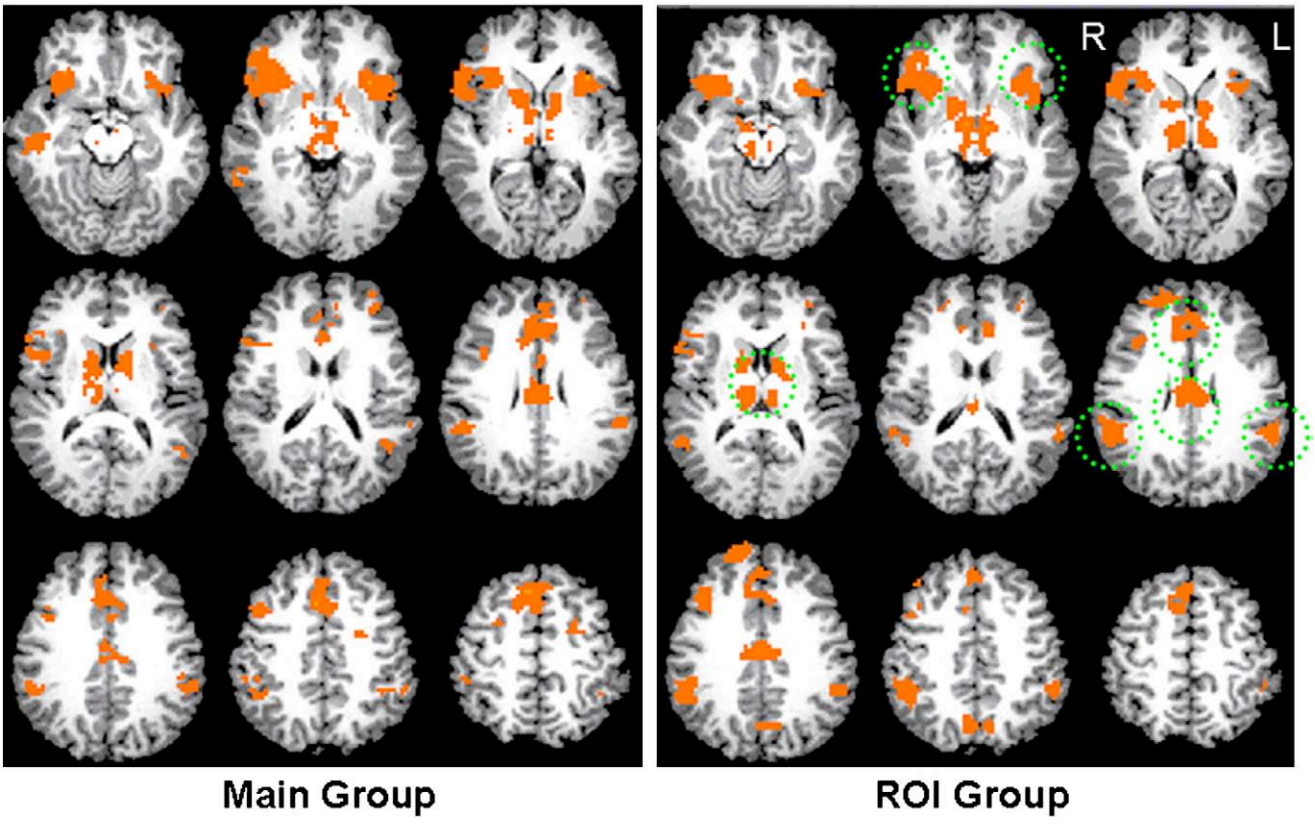

Fig. 3. Independent definition of ROIs and ROI analyses. Seven of the clusters found in the ROI group ($n = 12$), marked on the right panel with green circles, overlapped with the clusters found in the main group, shown on the left. Data are shown on nine horizontal slices in ascending order (the same normalized individual brain was used for the anatomical underlay in both cases). Note that the ROIs are marked on single slices for clarity but spanned several slices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Confirming countermeasure use

In the countermeasure condition we predicted an increased activation in the primary motor cortex representation of the left hand (right motor ROI) to the first two irrelevants, compared to the fourth one. A group ANOVA with factors of condition (no knowledge and countermeasure) and irrelevant (involving and not involving covert actions with left hand fingers) confirmed this prediction by showing an interaction between condition and irrelevant, $F(1,11) = 18.48$, $p < 0.001$, $\eta_p^2 = 0.63$. A similar result was found comparing the concealed knowledge and countermeasure conditions, $F(1,11) = 18.38$, $p < 0.001$, $\eta_p^2 = 0.63$, confirming that participants employed the countermeasure as instructed. Note that this result implies that information from the primary motor cortex could also be used to detect countermeasure use. However, this information may not be useful in the general case because it is only diagnostic of countermeasures involving a specific body part, and only in cases when the body part is known.

Discussion

This study shows that hemodynamic signals from lateral and medial prefrontal cortices could differentiate deceptive and honest responses but that such differential activation becomes much smaller when participants use a simple covert countermeasure. Critically, single subject classification accuracy, required for any deception test, is substantially reduced by the covert countermeasure, even in this controlled laboratory situation and with highly salient personal information. These effects are likely to be even stronger with incidentally acquired information, which is usually more difficult to detect with CIT paradigms (Rosenfeld et al., 2006, 2007).

This result is novel and important because this is the first fMRI study to apply countermeasures during a deception task. Only one fMRI study examined this issue non-experimentally by asking participants if they tried to beat the test, and no correlation was found between such attempts and correct classification (Kozel et al., 2005). However, it is unlikely that participants had practiced a countermeasure beforehand and tried to use it systematically, which would be the case for anybody seriously trying to beat the test.

We propose that the covert countermeasure used in this study was effective in large part because it assigned meaning and specific mental actions to the irrelevants, thus reducing the relative saliency of the probe within the stimulus sequence. According to this idea, the implicit motor component of this countermeasure may not be important for its efficacy: the same effect could be produced by performing other mental actions (e.g., recalling a certain episode from memory) that increase the relative saliency of the irrelevants. This saliency explanation is consistent with numerous findings. First, the activation in both lateral and medial prefrontal regions depends strongly on relative frequency and familiarity of the stimuli (Downar et al., 2002; Jones et al., 2002; Michelon et al., 2003), which are factors that modulate saliency, accrued knowledge, and meaningfulness. Second, these types of countermeasures also lower single subject classification rates in ERP studies using similar CIT paradigms (Rosenfeld et al., 2004) by reducing the size of the P300 component elicited by probes, compared to the irrelevants. Third, single subject classification rates in ERP studies are also lowered by purely cognitive countermeasures without an implicit motor component (Rosenfeld and Labkovsky, 2010).

In real life, participants do not have access to the irrelevants used in the test beforehand, as they did in this study. However, motivated suspects may devise irrelevant items on their own that are close to those used during the test since they would know a crime scene best. Furthermore, they may be well-practiced at

Please cite this article as: Ganis, G., et al., Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.11.025

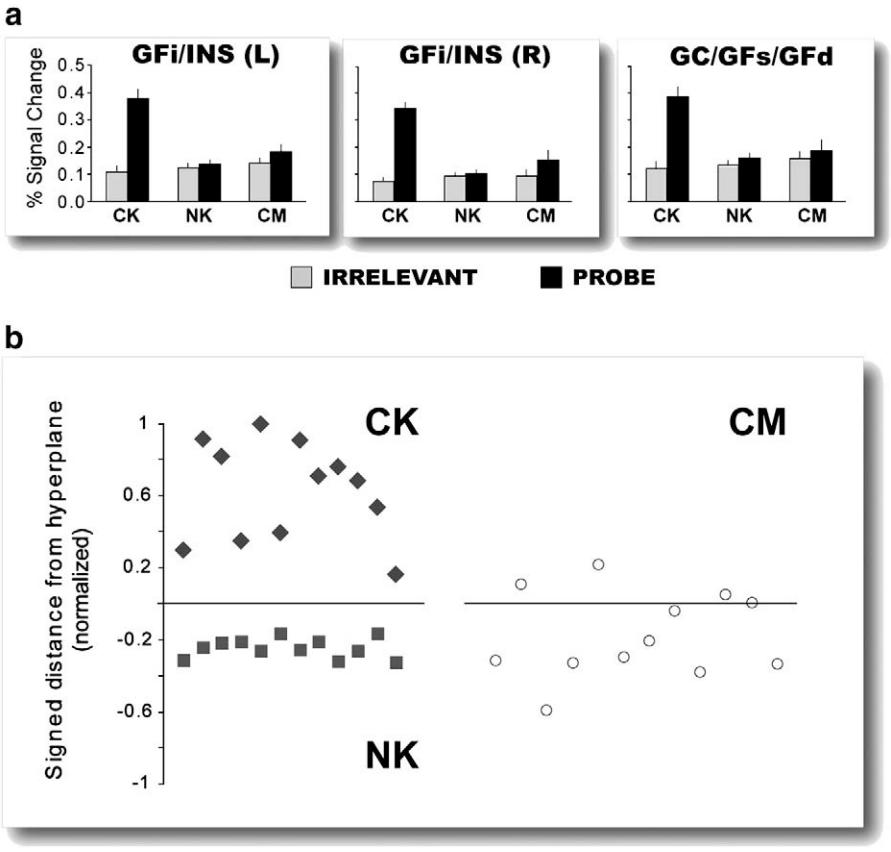

Fig. 4. Results of the single subject analyses: (a) bar graphs showing group activation to probes and irrelevants in the ROI triplet that best discriminated concealed and no knowledge cases (100% accuracy and largest margin): left GFi/INS, right GFi/INS, and GC/GFs/GFd. Error bars denote the standard error of the mean. Note that, given the differential response to targets required by the task, comparing brain activation between targets and irrelevants (or probes) is not informative. (b) Classification performance for all conditions achieved by the classifier trained to discriminate no knowledge and concealed knowledge cases, assessed with a jackknife procedure using the best ROI triplet. Each data point is a test case; the vertical axis shows the signed distance from the classification hyperplane, normalized by the maximum distance. Data were coded so that correctly classified no knowledge cases would have a negative signed distance, whereas correctly classified concealed knowledge and countermeasure cases would have a positive signed distance. All 12 concealed and no knowledge cases are classified correctly, but only 4 out of the 12 countermeasure cases are classified correctly.

methods to quickly associate new irrelevants with mental actions or memories (e.g., via imagery); such associations could be established during the first few trials and carried out consistently throughout the test.

Given that these countermeasures can be learned easily, this study provides evidence that additional research is needed before fMRI-based methods are sufficiently robust to detect concealed knowledge and deception accurately in the real world. For example, a scenario often described by companies selling lie detection services is one in which individuals accused of a crime may want to provide evidence of their innocence by undergoing an fMRI-based test. The results reported here indicate that finding no difference between the activation to probes and the irrelevants in a typical CIT paradigm does not imply that participants are honestly reporting ignorance about the probe; the result could instead be a false negative produced by covert countermeasures applied by individuals who have actually committed the crime under investigation. Although these results apply directly only to the specific laboratory paradigm used here (the 3-stimulus CIT protocol), they support the more general point that the vulnerability of the neuroimaging paradigms for deception detection to various countermeasures should be assessed and documented explicitly before they can be used in applied settings.

Competing interests statement

The authors declare that they have no competing financial interests.

Acknowledgment

This research was supported in part by the National Science Foundation (BCS0322611).

References

Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., Tsukada, M., Fujii, T., 2008. Neural correlates of true memory, false memory, and deception. Cereb. Cortex 18, 2811–2819.

Alkadhi, H., Crelier, G.R., Boendermaker, S.H., Golay, X., Hepp-Reymond, M.C., Kollias, S.S., 2002. Reproducibility of primary motor cortex somatotopy under controlled conditions. Am. J. Neuroradiol. 23, 1524–1532.

Ben-Shakhar, G., Elaad, E., 2003. The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. J. Appl. Psychol. 88, 131–151.

Bhatt, S., Mbwana, J., Adeyemo, A., Sawyer, A., Hailu, A., Vanmeter, J., 2008. Lying about facial recognition: an fMRI study. Brain Cogn. 69, 382–390.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

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.

Dale, A., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage 28, 663–668.

DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., Epstein, J.A., 1996. Lying in everyday life. J. Pers. Soc. Psychol. 70, 979–995.

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. J. Neurophysiol. 87, 615–620.

Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Capital City Press, Montpelier, Vermont.
Gamer, M., Bauermann, T., Stoeter, P., Vossel, G., 2007. Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. Hum. Brain Mapp. 28, 1287–1301.
Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P., Vossel, G., in press. fMRI-activation patterns in the detection of concealed information rely on memory-related effects. Soc. Cogn. Affect. Neurosci. doi:10.1093/scan/nsp005.
Ganis, G., Kosslyn, S.M., Stose, S., Thompson, W.L., Yurgelun-Todd, D.A., 2003. Neural correlates of different types of deception: an fMRI investigation. Cereb. Cortex 13, 830–836.
Ganis, G., Morris, R., Kosslyn, S.M., 2009. Neural processes underlying self- and other-related lies: an individual difference approach using fMRI. Soc. Neurosci. 4, 539–553.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15, 870–878.
Greely, H.T., Illes, J., 2007. Neuroscience-based lie detection: the urgent need for regulation. Am. J. Law Med. 33, 377–431.
Hand, D.J., Mannila, H., Smyth, P., 2001. Principles of Data Mining (Adaptive Computation and Machine Learning). MIT Press.
Honts, C.R., Devitt, M.K., Winbush, M., Kircher, J.C., 1996. Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. Psychophysiology 33, 84–92.
Jones, A.D., Cho, R.Y., Nystrom, L.E., Cohen, J.D., Braver, T.S., 2002. A computational model of anterior cingulate function in speeded response tasks: effects of frequency, sequence, and conflict. Cogn. Affect. Behav. Neurosci. 2, 300–317.
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. Biol. Psychiatry 58, 605–613.
Kozel, F.A., Padgett, T.M., George, M.S., 2004. A replication study of the neural correlates of deception. Behav. Neurosci. 118, 852–856.
Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540.
Langleben, D.D., Loughead, J.W., Bilker, W.B., Ruparel, K., Childress, A.R., Busch, S.I., Gur, R.C., 2005. Telling truth from lie in individual subjects with fast event-related fMRI. Hum. Brain Mapp. 26, 262–272.
Langleben, D.D., Schroeder, L., Maldjian, J.A., Gur, R.C., McDonald, S., Ragland, J.D., O'Brien, C.P., Childress, A.R., 2002. Brain Activity during simulated deception: an event-related finctional magnetic resonance study. Neuroimage 15, 727–732.
Lee, T.M.C., Au, R.K., Liu, H.L., Ting, K.H., Huang, C.M., Chan, C.C., 2009. Are errors differentiable from deceptive responses when feigning memory impairment? An fMRI study. Brain Cogn. 69, 406–412.
Lee, T.M.C., Liu, H.L., Chan, C.C., Ng, Y.B., Fox, P.T., Gao, J.H., 2005. Neural correlates of feigned memory impairment. Neuroimage 28, 305–313.
Lykken, D., 1974. Psychology and the lie detector industry. Am. Psychol. 29, 725–739.
Michelon, P., Snyder, A.Z., Buckner, R.L., McAvoy, M., Zacks, J.M., 2003. Neural correlates of incongruous visual information. An event-related fMRI study. Neuroimage 19, 1612–1626.
Mohamed, F.B., Faro, S.H., Gordon, N.J., Platek, S.M., Ahmad, H., Williams, J.M., 2006. Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation–initial experience. Radiology 238, 679–688.
Monteleone, G.T., Phan, K.L., Nusbaum, H.C., Fitzgerald, D., Irick, J.S., Fienberg, S.E., Cacioppo, J.T., 2008. Detection of deception using fMRI: better than chance, but well below perfection. Soc. Neurosci. 4, 528–538.
National Research Council, 2003. The Polygraph and Lie Detection.
Nature Neuroscience Editorial, 2008. Deceiving the law. Nat. Neurosci. 11, 1231.
Nose, I., Murai, J., Taira, M., 2009. Disclosing concealed information on the basis of cortical activations. Neuroimage 44, 1380–1386.
Nunez, J.M., Casey, B.J., Egner, T., Hare, T., Hirsch, J., 2005. Intentional false responding shares neural substrates with response conflict and cognitive control. Neuroimage 25, 267–277.
Phan, K.L., 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 Tesla. Acad. Radiol. 12, 164–172.
Rosenfeld, J.P., Biroschak, J.R., Furedy, J.J., 2006. P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. Int. J. Psychophysiol. 60, 251–259.
Rosenfeld, J.P., Labkovsky, E., Winograd, M., Lui, M.A., Vandenboom, C., Chedid, E., 2008. The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. Psychophysiology 45, 906–919.
Rosenfeld, J.P., Labkovsky, E., 2010. New P300-based protocol to detect concealed information: Resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. Psychophysiology 47, 1002–1010.
Rosenfeld, J.P., Soskins, M., Bosh, G., Ryan, A., 2004. Simple, effective countermeasures to P300-based tests of detection of concealed information. Psychophysiology 41, 205–219.
Rosenfeld, J.P., Shue, E., Singer, E., 2007. Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. Biol. Psychol. 74, 396–404.
Seymour, T.L., Seifert, C.M., Mosmann, A.M., Shafto, M.G., 2000. Using response time measures to assess guilty knowledge. J. Appl. Psychol. 85, 30–37.
Spence, S.A., Farrow, T.F., Herford, A.E., Wilkinson, I.D., Zheng, Y., Woodruff, P.W., 2001. Behavioural and functional anatomical correlates of deception in humans. NeuroReport 12, 2849–2853.
Spence, S.A., Kaylor-Hughes, C., Farrow, T.F., Wilkinson, I.D., 2008. Speaking of secrets and lies: the contribution of ventrolateral prefrontal cortex to vocal deception. Neuroimage 40, 1411–1418.
Vrij, A., 2008. Detecting Lies and Deceit, 2nd edition ed. Wiley, Chichester, UK.
Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E., Gray, J.R., 2008. The seductive allure of neuroscience explanations. J. Cogn. Neurosci. 20, 470–477.
Winograd, M.R., Rosenfeld, J.P., in press. Mock crime application of the Complex Trial Protocol (CTP) P300-based concealed information test. Psychophysiology. doi:10.1111/j.1469-8986.2010.01054.x.

Please cite this article as: Ganis, G., et al., Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.11.025