Detection of deception using fMRI: Better than chance, but well below perfection

Article *in* Social neuroscience · December 2009

DOI: 10.1080/17470910801903530 · Source: PubMed

Social Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t741771143

Detection of deception using fMRI: Better than chance, but well below perfection

George T. Monteleone [a]; K. Luan Phan [a]; Howard C. Nusbaum [a]; Daniel Fitzgerald [a]; John-Stockton Irick [a]; Stephen E. Fienberg [b]; John T. Cacioppo [a]
[a] University of Chicago, Chicago, IL, USA [b] Carnegie-Mellon University, Pittsburgh, PA, USA

First Published:December2009

To cite this Article Monteleone, George T., Phan, K. Luan, Nusbaum, Howard C., Fitzgerald, Daniel, Irick, John-Stockton, Fienberg, Stephen E. and Cacioppo, John T.(2009)'Detection of deception using fMRI: Better than chance, but well below perfection',Social Neuroscience,4:6,528 — 538

To link to this Article: DOI: 10.1080/17470910801903530

URL: http://dx.doi.org/10.1080/17470910801903530

SOCIAL NEUROSCIENCE, 2009, 4 (6), 528–538

Detection of deception using fMRI: Better than chance, but well below perfection

George T. Monteleone, K. Luan Phan, Howard C. Nusbaum, Daniel Fitzgerald, and John-Stockton Irick
University of Chicago, Chicago, IL, USA

Stephen E. Fienberg
Carnegie-Mellon University, Pittsburgh, PA, USA

John T. Cacioppo
University of Chicago, Chicago, IL, USA

Functional brain imaging has been considered a new and better technique for the detection of deception. The reasoning is that there is a neural locus or circuit for lying that is sensitive, specific, generalizable across individuals and measurement contexts, and robust to countermeasures. To determine the extent to which the group results predicted lying at the level of the individual, we reanalyzed data on 14 participants from a study that had previously identified regions involved in lying (thus satisfying the criterion for sensitivity). We assessed the efficacy of functionally determined brain regions based on the lie–truth contrast for $N-1$ participants to detect deception in the Nth individual. Results showed that no region could be used to correctly detect deception across all individuals. The best results were obtained for medial prefrontal cortex (mPFC), correctly identifying 71% of participants as lying with no false alarms. Lowering the threshold for a response increased hits and false alarms. The results suggest that although brain imaging is a more direct index of cognition than the traditional polygraph, it is subject to many of the same caveats and thus neuroimaging does not appear to reveal processes that are necessarily unique to deception.

Historian Ken Alder (2007) has described the polygraph and lie detection as an "American obsession," a pursuit steeped more in an American cultural fantasy than reliable evidence. Traditional polygraphy, using an array of physiological measures such as heart-rate, blood pressure, and electrodermal response, faced several fundamental problems: the sensitivity of the test, the effectiveness of the test in measuring what it purports to measure (specificity), the generalizability of the test across measurement contexts, and its robustness to countermeasures. In 2003, the Committee to Review the Scientific Evidence on the Polygraph concluded:

> Notwithstanding the limitations of the quality of the empirical research and the limited ability to generalize to real-world settings, we conclude that in populations of examinees such as those represented in the polygraph research literature, untrained in countermeasures, specific-incident polygraph tests can discriminate

Correspondence should be addressed to: George T. Monteleone, Center for Cognitive and Social Neuroscience, University of Chicago, 5848 S. University Avenue, Chicago, IL 60637, USA

We thank Alex Japour for his contributions to this research. Support for this research was provided by the Department of Army Research Grant #DAAD19-03-1-0042 and the MacArthur Law & Neuroscience Project.

www.psypress.com/socialneuroscience DOI:10.1080/17470910801903530

> lying from truth telling at rates well above chance, though well below perfection. (Committee to Review the Scientific Evidence on the Polygraph, 2003, p. 4)

New areas of research have since taken on the challenge of detecting deception by moving to neurophysiological measures, including the physiological detection of deception using functional magnetic resonance imaging (fMRI). The notion is that by imaging the brain processes involved in deception, the problems of traditional polygraphy can be avoided.

Several fMRI studies of lie detection have been conducted using binary-choice paradigms and nomothetic (group-level) analyses to determine whether it was possible to discriminate lying and truthful responding. Spence et al. (2001) reported a study in which participants either lied or told the truth about acts they had performed during the course of that day. A cue presented along with the question signaled to the participant whether to lie or tell the truth. Blood-oxygen-level dependent (BOLD) contrasts between lie and truth conditions revealed significantly greater activation during lying in the bilateral ventral prefrontal cortex (VLPFC), dorsolateral prefrontal cortex (DLPFC), medial prefrontal cortex (mPFC), and the left inferior parietal cortex. Spence et al. (2001) interpreted the results in terms of learning and motor response inhibition. Using an in-scanner adaptation of the Guilty Knowledge Test (GKT) with playing cards as stimuli, Langleben et al. (2002) instructed participants to either lie or tell the truth about cards they had in their possession. Results revealed a greater activation in the DLPFC and in the medial frontal gyrus extending to the anterior cingulate cortex (ACC) during lying than truth-telling. Also using the GKT, Phan et al. (2005) found lying to be associated with greater activity in the mPFC, bilateral DLPFC and VLPFC, and bilateral superior temporal sulci (STS). Activation in the mPFC and DLPFC were activated differentially by lying and truth-telling in all three studies.

Brain imaging has also been performed using experimental paradigms to simulate more realistic lying situations. Lee et al. (2002) used a forced-choice memory paradigm in which participants feigned memory impairments on two memory tests (digit memory task and autobiographical memory task) in a simulated assessment of malingering. The participants were requested to devise their own mental strategies for the lie, thereby increasing the external validity of the manipulation of lying. The activation for feigned memory impairment versus accurate recall conditions was similar for the two memory tasks and included anterior frontal regions (BA 9 and 10), bilateral parietal and temporal cortex, and subcortical regions including the caudate. The prefrontal region result suggested a role of information manipulation and executive control, with anterior regions (BA 10) participating in simultaneous processing of primary and secondary goals, and the DLPFC region processing anticipation of performance and working memory. The authors suggested that the activation of the parietal region was part of a network involved in calculated responses during lying, the temporal regions as participating in the cognitive manipulation of visual information, and the subcortical regions as involved in self-monitoring and inhibition of learned rules. A study by Nuñez, Casey, Egner, Hare, and Hirsch (2005) investigated the neural correlates of false versus true responses to autobiographical or non-autobiographical information. The investigators reported a contrast effect for autobiographical items in the mPFC, DLPFC, VLPFC, ACC, BA 9/10, and caudate. At the chosen levels of significance, however, no differences were found between lying and truth for non-autobiographical items, suggesting differences in neural response for lying depending on the content being tested. In an experimental study to assess the effects of the type of lie told, Ganis, Kosslyn, Stose, Thompson, and Yurgelun-Todd (2003) investigated lies that were rehearsed and spontaneous lies. Lies generally were associated with greater activation in BA 10, fusiform gyrus, and visual cortex whereas activation for lies was greater than truth-telling in mPFC and DLPFC only for spontaneous lies.

Across these studies, activity in the mPFC and DLPFC was greater for lying than truthful responding in all except the conditions in Ganis et al. (2003) in which the lies were autobiographical and well-rehearsed. It is not clear whether these regions are indicative of lying or are indicative of the increased self-monitoring and load on working memory that are often involved when one is lying. These studies have also focused on nomothetic analyses of the BOLD response as a function of lying versus truth-telling, where lying and truth-telling are not only known but are controlled by the experimenter. The critical question, which has limited the scientific validity

of the traditional polygraph, is whether lying can be discriminated from truth-telling on an individual level based on the observed physiological response when one is answering the critical question.

The first fMRI study endeavoring to discriminate lying from truth-telling at the idiographic level was reported by Kozel et al. (2005). BOLD responses were used to develop a mathematical model weighting regions of interest (ROIs) to distinguish lying from truth-telling, and a second group's BOLD responses to the task were used to classify them as lying or truth-telling. The authors reported classification at 90% accuracy using a combination of analysis in the middle frontal gyrus, inferior frontal gyrus, and anterior cingulate. This is only slightly better than the classification accuracy found for traditional polygraphy. No single region showed classification at better than 85% accuracy, and each cluster had a false alarm rate of nearly 20%. The combined analysis yielded false alarms for the 10% who were not successfully classified.

In a similar analysis based on a set of group predictor clusters, Langleben et al. (2005) reported single-cluster predictions at no better than 75% correct classification using the area under the curve of the receiver operator characteristic function, which plots probability of a false positive (specificity) against probability of correctly predicting a lie (sensitivity.) The area under the curve was described as the probability of accurately classifying a pair of observations, with the best single region (left inferior parietal lobule) showing an area under the curve of 75.1%. Using a stepwise logistic regression model across 14 ROIs, the authors reported a final model area under the curve of 84.7%.

These initial attempts at individual classification of lies using fMRI were comparable to the results of the traditional polygraph: well above chance, but well below perfection (Committee to Review the Scientific Evidence on the Polygraph, 2003). Indeed, the value of the area under the ROC curve reported by Langleben et al. (2005) lies in the middle of the accuracy results from traditional polygraphy summarized in the report of the Committee to Review the Scientific Evidence on the Polygraph (2003).

More recently, Spence, Kaylor-Hughes, Brook, Lankappa, and Wilkinson (in press) scanned a woman who had been convicted of poisoning a child but who continued to profess her innocence. Results revealed longer response latencies and greater activation of the VLPFC and ACC when she endorsed her accuser's version of events than when she endorsed her version of events; the authors concluded that: ''While we have not 'proven' that this subject is innocent, we demonstrate that her behavioural and functional anatomical parameters behave as if she were.'' Such a conclusion based on an individual's responses in an fMRI study can be questioned on logical grounds because lying is not the only cognitive process that has been associated with increased activation of the VLPFC and ACC, and little is known about how general the association is between lying and activation in these regions.

The purpose of the current study is to examine data from a typical fMRI study of lying (Phan et al., 2005) to determine how well lying could be differentiated from truth-telling at the idiographic level, with an emphasis on sensitivity, specificity, and generalizability across individuals. For instance, because the decision threshold can affect the sensitivity of measurement and the classification results, we used both a liberal and a conservative threshold for identifying significant differences in the BOLD response. In addition, prior studies examining idiographic classifications have restricted their analyses to the predictor clusters that significantly differentiated lying from truth-telling in the prediction sample. Such a procedure does not discriminate between the individual whose scan shows only that the ROIs identified in the predictor clusters were more active during lying than truth-telling—as set forth by the predictor model—and the individual whose fMRI scan shows that many if not most areas of the brain are more active during lying than truth-telling. Whether the inclusion of pseudo-ROIs increases or decreases specificity has not been examined previously. Consequently, we also performed analyses to determine the effects of the application of an ROI control region.

METHODS

A detailed account of experimental methods can be found in Phan et al (2005). Fourteen participants were given a modified version of the GKT in which they were asked to either tell the truth or lie about the possession of a playing card (e.g., 2 of hearts or 5 of clubs). Data were collected on a 4T Siemens platform, and preprocessed using Statistical Parametric Mapping (SPM), including slice-time correction, realignment, spatial

normalization to the Montreal Neurological Institute (MNI) template, resampling of functional images to 2-mm isotropic voxels, and spatial smoothing with a 6 mm full-width–half maximum Gaussian kernel. Preprocessed data were converted from ANALYZE (spm99) to a format for use with the analysis software AFNI (Cox, 1996). A canonical hemodynamic response function was convolved with the experimental conditions using the AFNI tool WAVER, and this model was regressed against the experimental data at each voxel to provide within-subjects statistical maps of the responses for each condition, as well as a map of the lie vs. truth contrast.

Nomothetic analyses

A group response map was generated by carrying out a whole-brain voxelwise one-sample *t*-test of the Lie > Truth contrast value across all 14 participants at $p < .01$ with a cluster size extent threshold of 119 contiguous voxels based on corner-to-corner connectivity. Phan et al. (2005) reported results only in regions where they had *a priori* hypotheses, namely ACC, mPFC, DLPFC, and VLPFC. In this case, a whole-brain voxelwise result was generated to identify all possible regions that differentiated lies from true statements.

Idiographic analyses

Idiographic analyses were carried out at two significance levels to assess the accurate detection of lies on an individual by individual basis using a conservative ($p < .01$) and a liberal ($p < .05$) criterion for Type I errors. To develop the predictive model for the idiographic analysis, the contrast images for 13 of the 14 participants were entered into the second stage of a random-effects analysis (one-sample *t*-test, two-tailed; $t = 3.057$, $\mathrm{df} = 12$, $p < .01$; $t = 2.179$, $\mathrm{df} = 12$, $p < .05$). This process was performed for each of the 14 participants, resulting in 14 group response maps, each of which was then used as the template for the neural signature of lying when examining each individual contrast map. That is, after group analysis from 13 participants yielded predictor clusters, the remaining 14th participant's data was thresholded at the same individual voxel extent levels as the group data, yielding a set of target clusters for the individual. Specifically, each remaining individual's contrast map was submitted to the same threshold based on the coefficient of the least-squares estimate of the empirical data to the model ($p < .01$, $t = 2.585$ and $p < .05$, $t = 1.964$). Significant regions were determined by applying an individual voxel probability threshold of $p < .01$ with a cluster volume of 952 μl (119 contiguous voxels) and again with a voxelwise threshold of $p < .05$ with a cluster volume of 5160 μl (645 voxels) based on corner-to-corner connectivity in 3D space, which was equivalent to a connectivity radius of 3.46 mm. The cluster volume was chosen as the means to correct for multiple comparisons ($\alpha < .05$). Cluster volume threshold was determined with a Monte Carlo simulation for which the input parameters modeled the analysis (voxel size $2 \times 2 \times 2$ mm, connectivity radius 3.46 mm, individual voxel $p = .01$ or $p < .05$, Gaussian FWHM filter width = 6 mm) executed within a mask of the entire brain (231,766 voxels) for 5000 iterations using the AFNI program AlphaSim. The Monte Carlo simulation randomly simulates "active" voxels within the mask according the probability and spatial parameters for the specified number of iterations, ultimately calculating the probability that a cluster of size *X* would occur by chance. The volume *X* is then used as a selection criterion on the experimental data to obtain activity clusters that meet the corrected α level (Forman et al., 1995; Xiong, Gao, Lancaster, and Fox, 1995).

For each participant, individual target cluster maps were overlaid on group predictor maps to assess signal detection. Significant voxels from the individual maps that fell within the nomothetic predictor map were tallied as hits if the valences from individual results matched those of the group result. No predictor maps yielded significant True > Lie clusters; however, some target maps yielded True > Lie clusters. If a True > Lie target cluster overlapped with a Lie > True predictor cluster, it was counted as a false alarm. If no individual results were found for a predictor cluster, it was tallied as a miss for that cluster. If hit and false-alarm target voxels were detected in the same predictor cluster, the cluster was classified according to the result that had the larger volume.

A threshold was set for the size of clusters that overlapped between the nomothetic predictor map and individual signal map for each subject. Overlap cluster size was determined by executing a Monte Carlo simulation within the mask of the predictor map voxels at the appropriate voxelwise threshold ($p < .01$ or $p < .05$) to determine the

cluster size corresponding to $\alpha < .05$. The cluster size for overlap regions at the voxelwise $p < .01$ level was 37 contiguous voxels, and at $p < .05$ it was 203 voxels.

The pseudo-ROI specificity control

Finally, we sought to discriminate between the individuals whose fMRI scan shows *only* that the predictor clusters were more active during lying than truth-telling and those who showed greater activation during lying than truth-telling across many areas, among them one or more of those specified in the predictor clusters. To do so, we formed a set of pseudo-ROIs that served as control regions for comparison with the predictor clusters. Assuming that the predictor clusters are in fact part of a reliable neural circuit, hit rates inside predictor clusters were hypothesized to be significantly higher than hit rates in the randomly sampled pseudo-ROIs. Given that results at the liberal threshold of $p < .05$ were expected to contain a considerably higher rate of false alarms than at $p < .01$, the $p < .05$ results were examined to determine whether the rate of false alarms would be decreased.

To form the pseudo-ROIs, the predictor ROIs were subtracted from a Talairach atlas map of all cortical grey matter, and for each predictor ROI a pseudo-ROI of equivalent volume was constructed from randomly sampled voxels in the remaining regions. Pseudo-ROIs were tested for signal detection in the same manner as group predictor ROIs, with a tally of significant voxels found at the individual level. A chi-square test (positive/negative × pseudo-ROI/actual-ROI) was used to compare frequency distributions of hits and false alarms in the pseudo-ROIs and predictor ROIs. Actual ROIs were expected to have a significantly higher hit rate than pseudo-ROIs, and predictor clusters that did not meet this criterion at $p < .05$ were discarded on grounds of non-specificity.

RESULTS

Nomothetic analyses

Whole-brain voxelwise group analysis showed that group Lie > Truth contrasts revealed significant activation in the medial prefrontal cortex (mPFC) in the region of the superior medial frontal gyrus, L inferior frontal gyrus (VLPFC), anterior cingulate, bilateral Brodmann's areas 9 and 10 (frontopolar region), bilateral temporo-parietal junction (TPJ), bilateral middle frontal gyrus (DLPFC), bilateral middle temporal gyurs, bilateral caudate, bilateral lingual gyrus, and the medial cingulate gyrus at a voxelwise $p < .01$ with a cluster minimum of 952 μl (119 contiguous voxels). Not all subjects' group analyses yielded the same set of regions; however, the majority of these regions (mPFC, VLPFC, ACC, BA 9.10, TPJ, R middle temporal, and L caudate) were found in all 14 group template maps. Results from a representative participant are illustrated in Figure 1 and detailed in Table 1.

Idiographic analyses

Figure 2 depicts the number of participants who showed activation in each ROI identified using data from the remaining 13 participants at two levels of detection threshold. First, more lenient detection thresholds are generally associated with more hits, fewer misses, and more false alarms. This was found in the present study: at $p < .01$, 43% of clusters were hits, 57% were misses, and 0% were false alarms, whereas at the more lenient detection threshold of $p < .05$, 53% were hits, 43% misses, and 4% of detected signals were false alarms. As would be expected from signal detection theory, decreasing the decision threshold increased the hit rate at the expense of also increasing the false alarm rate. Finally, no contrast was found in which truthful responses produced greater activation than lying.

The best classification at $p < .01$, with a classification success of 71%, was found in the mPFC in the region of the medial frontal gyrus (see Figure 2, top panel). No other regions predicted more than 7 of 14 participants. Four out of five subjects were classified in L Middle Temporal gyrus, but this region achieved significance and, thus, served as an ROI, in only 36% of the predictor maps. No false positives were observed in any of these regions.

Lowering the detection threshold to $p < .05$ increased the number of template maps in which the areas achieved significance and the percentage of correct classifications in some regions. For instance, 78% of participants were classified successfully as lying in mPFC and BA 9/10, and 64% in VLPFC, ACC, and Left TPJ (see Figure 2, middle panel). However, these

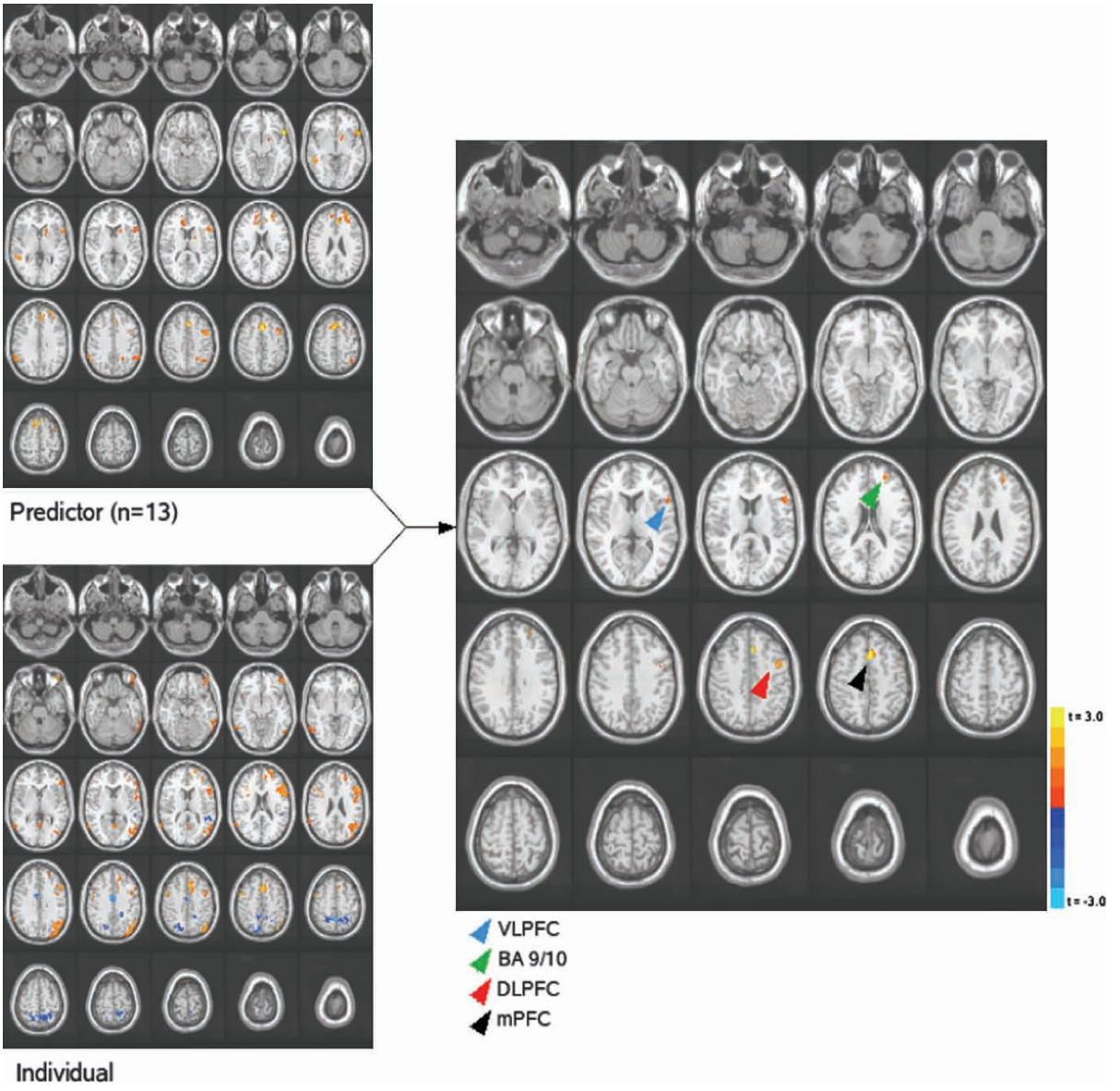

Figure 1. An illustration of the signal detection process from a representative subject. The Lie >True group predictor map (top left) was determined using 13 subjects' data, and was used to mask the individual Lie >True map (bottom left). Both maps were subjected to a voxelwise threshold of $p < .01$ with a minimum cluster size of 119 contiguous voxels as determined by a whole-brain Monte Carlo simulation at $\alpha < .05$. The overlap map was cluster-limited by another Monte Carlo simulation carried out in the predictor map, resulting in a cluster minimum of 37 voxels.

increased levels of detection came at the cost of increased false positives.

The effects of applying a pseudo-ROI control are illustrated in the bottom panel of Figure 2. If activation in a predicted ROI was accompanied by activation in the matched pseudo-ROI—that is, if the activation was nonspecific—then activation in the predicted ROI was treated as nondiagnostic. The results from four participants were affected by the implementation of pseudo-ROIs. The regions excluded were not entirely consistent across the participants, however. For participant #8, mPFC, BA 9/10, left TPJ, and left caudate regions were excluded, and for participant #14, right lingual gyrus was excluded. All five exclusions would otherwise have been considered hits.

DISCUSSION

Recent reports tout fMRI lie-detection as a means to spot liars (Peplow, 2004), detect terrorists (Wild,

TABLE 1

Lie > Truth regions for the representative subject pictured in Figure 1: The top table shows the predictor ROIs; the bottom table shows detected overlap regions

	Volume (μl)	*Hemisphere*	*Region(s)*	*TT atlas region(s)*	*BA*	*X*	*Y*	*Z*	*Mean B-value*
1	9104	R/L R/L L	mPFC BA 9,10 Anterior cingulate	Superior frontal gyrus Medial frontal gyrus Anterior cingulate	6, 9, 10, 32	4	47	26	.904
2	3544	L	VLPFC	Inferior frontal gyrus	47	−53	37	−18	.816
3	2552	L	Temporoparietal	Inferior parietal lobule	13, 40	−42	−36	26	.531
4	2120	R R	BA 9/10 Anterior cingulate	Medial frontal gyrus Anterior cingulate	10	14	61	2	.503
5	2040	L	DLPFC	Middle frontal gyrus	6	−42	23	31	.654
6	1664	L	Caudate	Caudate	25, 47	−15	29	−14	.548
7	1520	R	Temporoparietal	Superior temporal gyrus	42	60	−33	17	.660
8	1296	R	Middle temporal	Middle temporal gyrus	20	53	−25	−18	.855
1	1120	R/L	mPFC	Medial frontal gyrus	32	2	37	32	—
2	728	L	VLPFC	Inferior frontal gyrus	47	−53	37	4	—
3	672	L	BA 9/10	Superior frontal gyrus	9, 10	−24	68	7	—
4	480	L	DLPFC	Middle frontal gyrus	6	−22	27	38	—

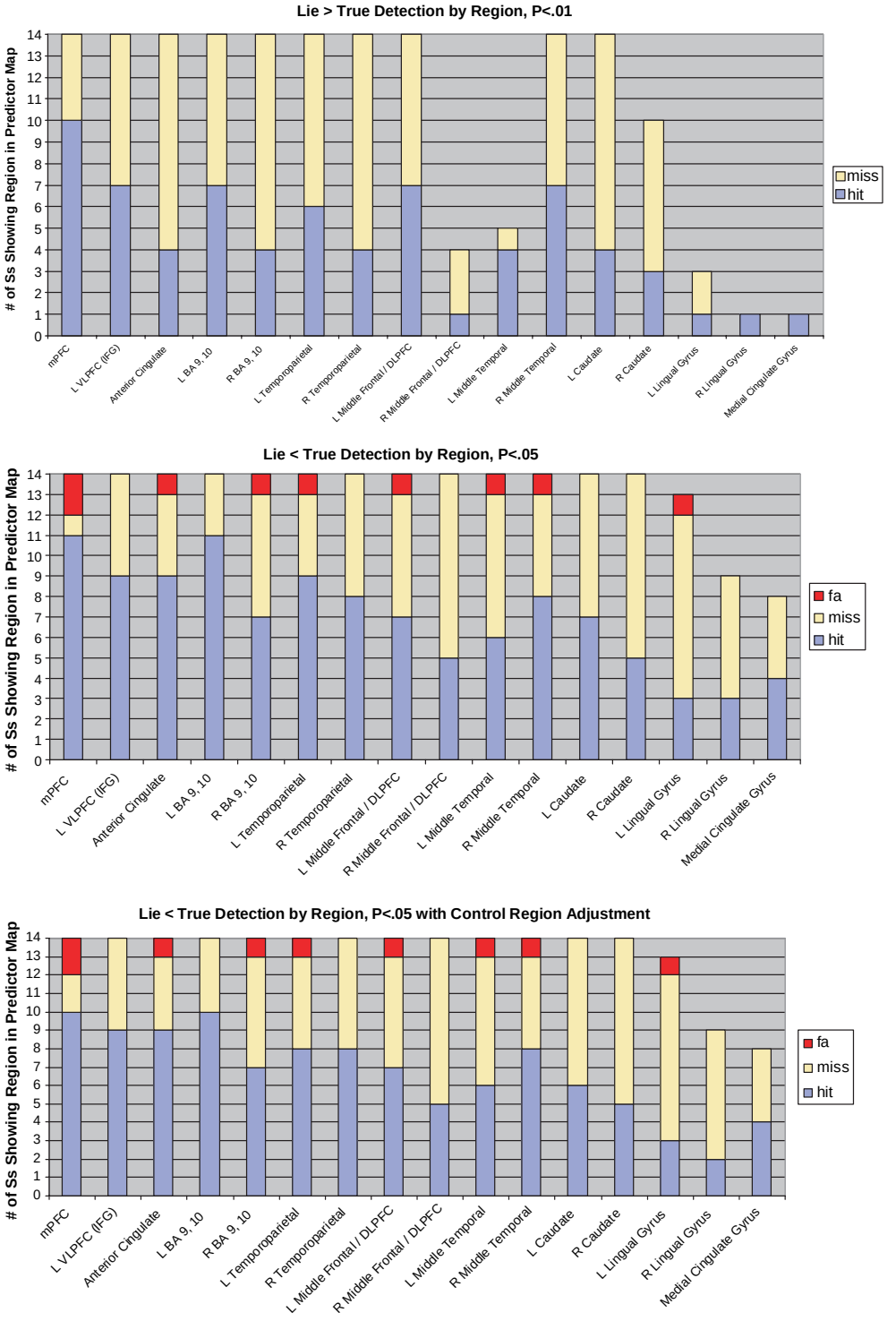

Figure 2. (Top panel) The number of hits and misses in each ROI at a significance threshold of $p < .01$ with a minimum overlap cluster size of 37 voxels. The total bar height indicates the number of subjects for which a response was predicted in each region. Some regions (R DLPFC, L MTG, R caudate, L and R lingual gyri, and medial cingulate) were predicted for only a fraction of subjects. No false alarms (True >Lie in a region predicting Lie >True) were detected at this significance level. Some regions (R DLPFC, L middle temporal, R caudate, L and R lingual, and cingulate gyrus) were significantly active in only a portion of subjects' predictor maps. (Middle panel) The number of hits and misses in each ROI at a significance threshold of $p < .05$ with a minimum overlap cluster size of 203 voxels. (Bottom panel) The number of hits, misses, and false alarms in each ROI at a significance threshold of $p < .05$ after the application of the pseudo-ROI control region (see text for an explanation of pseudo-ROIs).

2005), and reveal secret intentions in the brain (Haynes et al., 2007). Efforts to privatize new deception detection technologies have brought forth ethical concerns (Pearson, 2006), issues of personal privacy (Olson, 2005) and legal concerns about fourth amendment rights and personal liberty (Boire, 2005; Greely & Illes, 2007). Forensic applications and reports of effective ''brain fingerprinting'' (Farwell & Smith, 2001) and fMRI (Spence et al., in press) have been accompanied by efforts on the part of private interests to apply brain imaging for commercial gain. The present results raise several questions about the claim that there is an invariant neural locus unique to lying that is sensitive, specific, and consistent across individuals.

First, using a variant of the GKT, lying and truth-telling in the current study were associated with differential activation in as many as 16 different brain regions in some template maps. The classification of an individual as lying or not based on ROIs specified by the template maps proved to be well above chance, but no better than the levels typically observed using traditional polygraphy (Committee to Review the Scientific Evidence on the Polygraph, 2003). The best classification accuracy found for any ROI using a stringent classification threshold was 71% (with no false positives), found using the mPFC. Using a less stringent criterion for identifying areas of brain activation led to more hits, with activation in two ROIs in the frontal cortex producing correct classifications in 78% of cases, but false-alarm rates also increased across the areas that were identified.

Even though activation of the mPFC provided better than chance classification of individuals as lying in the present study, it does not mean that we detected a lie response *per se*. The mPFC has been associated with various cognitive processes including self-awareness and self-referent processing, mentalizing, executive functioning, and theory of mind (e.g., Frith & Frith, 2003; Krendl & Heatherton, in press; Saxe, Carey, & Kanwisher, 2004). Accurate classifications in the current study could have resulted from one or more of these cognitive processes occurring more often in the lying than in the truthful response conditions. It is obvious that knowing that lying leads to a skin conductance response does not logically mean that a skin conductance response indicates a person is lying, because lying is not the only process that produces a skin conductance response. Similarly, knowledge that lying leads to regional brain activation (e.g., mPFC) does not imply logically that activation of that region marks a person who is lying.

Brain imaging is a more direct index of cognition than the traditional polygraph, but it may be subject to many of the same caveats. For instance, we found that adopting a more lenient threshold increased hits, but at the expense of increasing false positives, just as has been found in studies of traditional polygraphy. False positives can be particularly pernicious in legal and employment applications, and even low false positive rates can lead to large numbers of truthful individuals being classified as lying when the base rate for lying is low. For instance, if the fMRI detection procedure has an accuracy index of .90 in a hypothetical population of 10,000 examinees that includes 10 liars, a detection threshold set to detect 80% of the liars will miss 2 of the 10 liars and will falsely classify 1,598 truth-telling individuals as lying (Committee to Review the Scientific Evidence on the Polygraph, 2003).

Second, our experimental paradigm made it possible to know when participants were telling the truth or lying, thereby making it possible to determine regions of brain activation that differentiated lying and truth-telling (i.e., a template). Typically, it is not possible to know whether a participant is telling the truth or lying. The intersubject variability in brain response to lying observed in the current study means that the template identified using a nomothetic approach may fail to characterize accurately many individuals' brain response to lying. This was evident in the present study, where nomothetic maps permitted lies to be accurately classified in the case of most but not all individuals. Templates designed at the idiographic rather than the nomothetic level may be worth investigating to discriminate lying from truth-telling, but the merit of such an approach has yet to be proven. To develop such a template, for instance, individuals would need to lie and to tell the truth in known sequences in response to multiple questions. This procedure may be limited if participants are not cooperative or if different kinds of lies are associated with different patterns of brain activation.

Third, the empirical support to date has raised questions about the notion that there is an invariant neural signature for lying. The card version of the GKT is a simple task that may limit individual variability in type of lie that is

emitted. Although a card-based guilty knowledge test serves as a model for lying, it may be an overly simplified paradigm for testing legitimate lie-detection, and does not address the issue of differences in response according to different types of lie or the type of information involved. Thus, any level of success of lie classification based on this paradigm is but an early step towards more ecologically valid scenarios. Indeed, the research by Ganis et al. (2003) suggests that different types of lies may be associated with completely different patterns of brain activation. This seems plausible, as different types of lies are the product of one or more distinguishable component processes, which should be reflected in differences in neural activation.

Fourth, the brain imaging research on lying to date has not considered what detection problems might be introduced by participants who use physical or mental countermeasures. If a participant were to perform serial subtraction or to focus on suppressing an irrelevant response when responding truthfully and lying, for instance, would the pattern of neural activation that differentiates lying from truth-telling still be evident? Or, in light of the nomothetic template found in the present study, would thinking about the mental state of the examiner and oneself as the object of interrogation when responding truthfully act as a cognitive countermeasure to the detection of deception? Such questions will be important to address before the applied value of lie detection using fMRI can be determined.

Finally, social, cultural, and linguistic differences across participants in the understanding of questions and the meaning and appropriateness of deception may contribute to differences in regional activation patterns observed during lying versus truth-telling. In the absence of an invariant neural signature for lying, successful detection of deception using fMRI may require a better appreciation of possible sociocultural effects.

In sum, fMRI as a method for the detection of deception should carry the same burden of proof as demanded of any other method. As in prior studies, fMRI analyses permitted the differentiation of deceptive and truthful responding at the group level. The classification of lying based on contrast maps, however, was better than chance but far from perfect. Two different reasons for misclassifications were identified. In some cases, the activation of a predicted ROI did not reflect specific activation in this region but instead reflected activation of a very large area that included part of the predicted ROI. The use of pseudo-ROIs proved useful in identifying such cases. The second and more common reason for misclassifications was that one or more of the regions found to differentiate lying from truthful responding in the template map were not associated with significant differences in activation in the test participant. Using more lenient thresholds for detecting activation only nominally improved classification accuracy and came at the expense of increased false alarm rates. Closer inspection of these individuals' contrast maps (lie – true) suggested that a few subjects showed lying associated with very different regions of activation than most individuals. Although this is speculative, lying for these individuals may have been achieved by equally atypical means. Together, these results suggest that, although fMRI may permit investigation of the neural correlates of lying, at the moment it does not appear to provide a very accurate marker of lying that can be generalized across individuals or even perhaps across types of lies by the same individuals.

REFERENCES

Alder, K. (2007). *The lie detectors: The history of an American obsession*. New York: The Free Press.

Blasi, G., Goldberg, T., Weickert, T., Das, S., Kohn, P., Zoltick, B., Bertolino, A., Callicott, J., Weinberger, D. R., & Mattay, V. S. (2006). Brain regions underlying response inhibition and interference monitoring and suppression. *European Journal of Neuroscience*, *23*(6), 1658–1664.

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *Journal of Neuroscience*, *26*(3), 916–925.

Boire, R. G. (2005). Searching the brain: the fourth amendment implications of brain-based deception detection devices. *American Journal of Bioethics*, *5*(2), 62–63.

Committee to Review the Scientific Evidence on the Polygraph (2003). *The polygraph and lie detection.* Washington, DC: National Academy Press.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.

Farwell, L. A., & Smith, S. S. (2001). Using brain MERMER testing to detect knowledge despite efforts to conceal. *Journal of Forensic Sciences*, *46*(1), 135–143.

Forman, S. D, Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a

cluster-size threshold. *Magnetic Resonance Medicine*, *33*, 636–647.

Frith, D. D., & Frith, U. (2003). Interacting minds: A biological basis. *Science*, *286*, 1692–1695.

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. *Cerebral Cortex*, *13*, 830–836.

Greely, H. T., & Illes, J. (2007). Neuroscience-based lie detection: the urgent need for regulation. *American Journal of Law & Medicine*, *33*, 377–431.

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, D. (2007). Reading hidden intentions in the human brain. *Current Biology*, *17*, 323–328.

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*(8), 605–613.

Kozel, F. A., Padgett, T. M., & George, M. S. (2004a). A replication study of the neural correlates of deception. *Behavioral Neuroscience*, *118*, 852–856.

Kozel, F. A., Revell, L. J., Lorberbaum, J. P., Shastri, A., Elhai, J. D., Horner, M. D., Smith, A., Nahas, Z., Bohning, D. E., & George, M. S. (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.

Krendl, A. C., & Heatherton, T. F. (in press). Components of the social brain. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences.* New York: Wiley.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Choldress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *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 functional magnetic resonance study. *NeuroImage*, *15*, 727–732.

Lee, T. M., Liu, H. L., Tan, L. H., Chan, C. C. H., Mahankali, S., Feng, C. M., Hou, J., Fox, P. T., & Gao, J. H. (2002). Lie detection by functional magnetic resonance imaging. *Human Brain Mapping*, *15*, 157–164.

Nuñez, 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.

Olson, S. (2005). Brain scans raise privacy concerns. *Science*, *307*, 1548–1550.

Pearson, H. (2006). Lure of lie detectors spooks ethicists. *Nature*, *441*, 918–919.

Peplow, M. (2004). Brain imaging could spot liars. *Nature News.* Retrieved January 15, 2008 from http://www.bioedonline.org/news/news.cfm?art = 1409

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. *Academic Radiology*, *12*, 164–172.

Rahm, B., Opwis, K., Kaller, C. P., Spreer, J., Schwarzwald, R., Seifritz, E., Halsband, U., & Unterrainer, J.M. (2006). Tracking the subprocesses of decision-based action in the human frontal lobes. *NeuroImage*, *30*(2), 656–667.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87–124.

Spence, S.A., Farrow, T.F., Herford, A.E., Wilkinson, I.D., Zheng, Y., & Woodruff, P.W. (2001). Behavioral and functional anatomical correlates of deception in humans. *NeuroReport*, *12*, 2849–2853.

Spence, S. A., Kaylor-Hughes, C. J., Brook, M. L., Lankappa, S. T., & Wilkinson, I. D. (in press). 'Munchausen's syndrome by proxy' or a 'miscarriage of justice'? An initial application of functional neuroimaging to the question of guilt or innocence. *European Psychiatry.*

Wild, J. (2005). Brain imaging ready to detect terrorists, say neuroscientists. *Nature*, *437*, 457.

Wirsing, B. (2007). *Revealing secret intentions in the brain (Press Release)*. Munich: Max Planck Society for the Advancement of Science.

Wolpe, P. R., Foster, K. R., & Langleben, D. D. (2005). Emerging neurotechnologies for lie detection: Promises and perils. *The American Journal of Bioethics*, *5*(2), 39–49.

Xiong, J., Gao, J.-H., Lancaster, J. L., & Fox, P. (1995). Clustered pixels analysis for functional MRI activation studies of the human brain. *Human Brain Mapping*, *3*, 287–301.