RAPID COMMUNICATION

Brain Activity during Simulated Deception: An Event-Related Functional Magnetic Resonance Study

D. D. Langleben,¹ L. Schroeder, J. A. Maldjian,* R. C. Gur, S. McDonald, J. D. Ragland, C. P. O'Brien, and A. R. Childress

Department of Psychiatry and *Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received May 30, 2001; published online January 4, 2002

INTRODUCTION

According to the traditional approach stemming from Saint Augustine, deception of another individual is intentional negation of subjective truth (Augustine, 1948). This concept suggests that inhibition of truthful response is a prerequisite of intentional deception. Deception has major legal, political, and business implications and there is a strong general interest in objective methods for its detection (Holden, 2001).

Multichannel physiological recording (polygraph) is currently the most widely used method for the detection of deception (Office of Technology Assessment, 1990). The effectiveness of the polygraph in the study and detection of deception is limited by reliance on peripheral manifestations of anxiety (skin conductance, heart rate, and respiration) (Office of Technology Assessment, 1983). Scalp-recorded event-related potentials (ERPs) have also been used in the study of deception. These series of voltage oscillations, which reflect the neuronal activity associated with a sensory, motor, or cognitive event, provide high temporal resolution, but their source in the brain cannot be uniquely localized (Hillyard and Anllo-Vento, 1998). The P-300 (P-3) wave of the ERP appears in response to rare, meaningful stimuli with a 300- to 1000-ms latency (Rosenfeld, 2001). Amplitude and latency of the P-3 have been associated with deception, suggesting that the cognitive differences between lying and telling the truth could be associated with changes in other correlates of brain activity, such as regional cerebral flow (rCBF) (Holden, 2001; Rosenfeld, 2001). Unlike the ERP, the spatial resolution of blood oxygenation level-dependent (BOLD) fMRI is sufficient to localize changes in rCBF related to regional neuronal activity during cognition (Chen, 1999). There have been no peer-reviewed reports on the use of fMRI to study deception.

The Guilty Knowledge Test (GKT) is a method of polygraph interrogation that facilitates psychophysiological detection of prior knowledge of crime details that would be known only to a suspect involved in the crime (Lykken, 1991; Elaad et al., 1992). The GKT has been adapted to model deception in psychophysiological (Furedy and Ben-Shakhar, 1991; Furedy et al., 1994; Elaad and Ben-Shakhar, 1997) and ERP research (Rosenfeld et al., 1988; Farwell and Donchin, 1991; Allen et al., 1992). In a typical laboratory GKT,
the subject is instructed to answer “No” in response to a series of questions or statements, the answer to some of which is known to be “Yes” to both the investigator and the participant; however, the participant may be unaware of investigator’s knowledge. An important distinction between the forensic and the laboratory GKT is that in the latter, the deception is endorsed by the investigator (Furedy and Ben-Shakhar, 1991). While still conforming to the Augustinian definition, such simulated deception may not be perceived by the participant as an immoral act and is less likely to invoke guilt or anxiety than the forensic version.

Since deception-induced mood and somatic states may vary across individuals, a search for a marker of deception independent of anxiety or guilt is justified. We hypothesized that regional brain activity elicited by the inhibition of the truth response during intentional deception could serve as such a marker. Our specific hypotheses were: (1) The difference in brain activity between lying and telling the truth on the GKT can be detected and localized with BOLD fMRI. (2) In normal adults, the GKT would activate parts of the cingulate and prefrontal cortex associated with response inhibition.

**METHODS**

Twenty-three healthy right-handed participants (11 men and 12 women) ages 22 to 50 years (average 32), education 12–20 years (average 16), were recruited from the University of Pennsylvania community. Participants were screened with Symptom Checklist-90—Revised (SCL-90-R) and a DSM-IV-based interview before the scan and questioned about symptoms of anxiety they had during and after the scan [(Derogatis et al., 1976) SCL-90-R items 2, 4, 12, 17, 23, 31, 39, 55, 57, 72, 78].
We adapted the “high-motivation” version of the GKT described by Furedy and Ben-Shakhar (1991) as follows: (1) Instead of handmade cards with written numbers, we used numbered playing cards (Fig. 1), (2) We added two non-salient card types to ensure alertness and attention and to control for the effect of repetition of the salient cards. The need for the multiple repetition of the salient stimuli and thus a special effort to maintain participants’ alertness was dictated by the event-related fMRI paradigm design (Aguirre, 1999). Four categories of cards were used: 5 of Clubs (“Lie”), 11 different numbered playing cards (“Non-Target”), 2 of Hearts (“Truth”), and 10 of Spades (“Control”). The Lie, Nontarget, and Truth cards carried the question: “Do you have this card?” The Control was accompanied by a question “Is this the 10 of spades?” The Control forced the participants to read the questions on top of all cards, rather than give an indiscriminate “No” response. The Non-Target introduced an appearance of randomness and reduced habituation and boredom that is expected if only three cards were repeatedly presented over 22 min. Truth was presented the same number of times as Lie to control for the effect of repetition (habituation).

Participants were told that if they lied about any card other than the one hidden in their pocket the reward would be forfeited. This amounted to endorsing the truth about not having the Non-Target and Truth cards, denying the truth (lying) about not having the Lie card, and endorsing the truth about the Control being the 10 of Spades. Lie, Truth, and Control were presented 16 times and each Non-Target was presented only twice, for a total of 88 stimuli. A random numbers generator was used to order the stimuli, which were presented for 3 s each. The interstimulus interval was 12 s (Aguirre, 1999), and thus the entire session lasted 1320 s (22 min).

PowerLab software (Chute, 1996) (MacLaboratory, Inc., Devon, PA) was used to assemble the GKT from scanned images of selected numbered playing cards and add-on graphics (Fig. 1). All participants were familiar with card games but had no history of problem gambling. Participants were asked to pick one of three sealed envelopes, all of which contained a $20 bill and a 5 of Clubs playing card. Participants did not know that all envelopes held the same contents. Participants were asked to secretly open the envelope, memorize the card, put it back in the envelope, and hide it in their pocket. Participants were told that they would be able to keep the $20 if they succeeded in concealing the identity of their card from a “computer” that would administer the GKT and analyze their brain activity during the MRI session. Participants were then positioned in a 4 Tesla MRI scanner (GE Signa), equipped for echo-planar imaging. An Apple computer running PowerLab and interfaced with a video projector was used to back-project the GKT onto a screen at the participants’ feet, visible through a mirror inside the radiofrequency head coil. “Yes” or “No” responses were made with a right-thumb press on a two-button fiber-optic response pad (Current Designs, Philadelphia, PA). Responses were fed back to the Apple computer and recorded by the PowerLab. Image acquisition was synchronized with stimuli presentation in an event-related fashion.

Sagittal T1-weighted localizer and a T1-weighted acquisition of the entire brain were performed in the axial plane (24 cm FOV, 256 × 256 matrix, 3-mm slice thickness). This sequence was used both for anatomic overlays of the functional data and spatial normalization of the data sets to a standard atlas. Functional imaging was performed in the axial plane using multislice gradient-echo echo-planar imaging with a field of view of 24 (frequency) × 15 (phase) and acquisition matrix of 64 × 40 (21 slices, 5 mm thickness, no skip, TR = 3000, TE = 40, and effective voxel resolution of 3.75 × 3.75 × 4 mm). The fMRI raw echo amplitudes were saved and transferred to a Sun Ultrasparc 10 (Sun Microsystems, Mountain View, CA) for offline reconstruction. Correction for image distortion and alternate k-space line errors on each image was based on the data acquired during phase-encoded reference imaging (Alsop, 1995).

Statistical analysis was performed with SPM99 (Wellcome Department of Cognitive Neurology, UK) (Friston et al., 1995a,b) implemented in Matlab (The Mathworks, Inc., Sherborn, MA), with an IDL (Research Systems, Inc., Boulder, CO) interface developed in-house. The T1-weighted images were normalized to a standard atlas (Talairach, 1988) within SPM99. Slice-acquisition timing correction was performed on the functional data using sync interpolation. Functional data sets were then motion corrected within SPM99 using the first image as the reference. Functional data sets were normalized to Talairach space using image header information to determine the 16-parameter affine transform between the data sets and the T1-weighted images (Maldjian et al., 1997), in combination with the transform computed within SPM99 for the T1-weighted anatomic images in Talairach space. The normalized data sets were resampled to 4 × 4 × 4 mm within Talairach space using sync interpolation. The data sets were smoothed using a 12 × 12 × 12-mm full width at half-maximum Gaussian smoothing kernel. For the statistical parametric mapping (SPM) analysis, a canonical hemodynamic response function with time and dispersion derivatives was employed as a basis function, with proportional scaling of the image means. Temporal smoothing, detrending, and high pass filtering were performed as part of the SPM analysis. SPM projection maps (SPMs) were generated using the general linear model (GLM) within SPM99. Within-subject contrasts between GLM regression coefficients were generated within SPM99 for the main contrast: “Lie vs Truth.” A second-level
analysis was performed to generate group SPMs using a random-effects model within SPM99 with the individual contrast maps (Holmes, 1988). The resulting SPM$_{t}$ map was transformed to the unit normal distribution SPM$_{Z}$ and thresholded at a $P_{0.01}$, corrected for spatial extent ($P_{0.05}$), using the theory of Gaussian fields as implemented in SPM99. Anatomic regions were automatically defined using a digital MRI atlas (Kikinis, 1996), which we had previously normalized to the same SPM99 Talairach template for use with our fMRI data. The resultant thresholded SPM was overlaid on a standard T1 template with MEDx software. Subjects were excluded from analysis if they made more than two errors responding to the Truth or Lie stimulus or more than three errors total on the GKT. Participants were also excluded from analysis if their individual $Z$ maps contained nonanatomical curvilinear change in $Z$ values, indicating a motion artifact (Hajnal et al., 1994). Montreal Neurological Institute coordinates (SPM99 output) were converted into stereotactic Talairach coordinates using a nonlinear transform (Duncan et al., 2000) and anatomical and Brodmann areas (BA) determined from the Talairach atlas (Talairach, 1988). Within SPM99, a “contrast” between condition A and condition B returns only positive differences (an increase); to detect a decrease a reversed subtraction (B minus A) is performed.

**RESULTS**

**Excluded Participants**

Four participants were excluded because of motion artifact (see Methods) and one—because of a 100% error rate on the GKT. The final number of participants included in the analysis was 18.

**GKT Performance**

Correct response rate was 97 to 100%. In a total of 88 trials, nine participants made no errors, four made one error, three made two errors, and two made three errors. None made more than two errors on the Lie, Truth, or Control cards.

**Imaging Data**

In the “Lie vs Truth” contrast (Table 1, Fig. 2), there are two clusters of significant BOLD signal increase. The first is a 146-voxel cluster extending from the left anterior cingulate gyrus (ACC) to the medial aspect of the right superior frontal gyrus (SFG), including BA 24, 32, and 8. The second is a 91-voxel cluster, U-shaped along the craniocaudal axis, extending from the border of the prefrontal to the dorsal premotor cortex (BA 6 bordering on BA 3 and 4) and also involving the anterior parietal cortex from the central sulcus to the lower bank of the intraparietal sulcus (BA 1–3 to the edge of BA 40). There were no regions with significant signal decrease.

**Anxiety**

All participants denied symptoms of anxiety during or after the scan.

**DISCUSSION**

ACC and dorsolateral prefrontal cortex (DLPFC) activation has been reported in executive function tasks involving inhibition of a “prepotent” (e.g., basic) response, divided attention, or novel and open-ended responses (Carter et al., 1998). Recent fMRI studies manipulating the Stroop task, a response inhibition paradigm, have narrowed the role of the ACC to monitoring the conflicting response tendencies and showed that the degree of right ACC activation is proportional to the degree of response conflict and inversely related to the left DLPFC activation (Carter et al., 2000; MacDonald et al., 2000). Increased activation of the right ACC but not the DLPFC during the Lie response suggests that a conflict with the prepotent response (Truth) and eventually its inhibition are taking place.

**TABLE 1**

<table>
<thead>
<tr>
<th>Cluster size (voxels)</th>
<th>Z</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>BA</th>
<th>Gyrus</th>
</tr>
</thead>
<tbody>
<tr>
<td>146</td>
<td>3.8</td>
<td>-1</td>
<td>16</td>
<td>29</td>
<td>24;32</td>
<td>Anterior cingulate</td>
</tr>
<tr>
<td></td>
<td>3.17</td>
<td>3</td>
<td>28</td>
<td>43</td>
<td>6;8</td>
<td>Right superior frontal</td>
</tr>
<tr>
<td></td>
<td>3.15</td>
<td>0</td>
<td>24</td>
<td>52</td>
<td>8</td>
<td>Superior frontal</td>
</tr>
<tr>
<td>91</td>
<td>3.58</td>
<td>-57</td>
<td>-23</td>
<td>41</td>
<td>1;2;3;40</td>
<td>Left postcentral</td>
</tr>
<tr>
<td></td>
<td>3.40</td>
<td>-54</td>
<td>-15</td>
<td>38</td>
<td>3;4;6</td>
<td>Left pre- and postcentral</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>-50</td>
<td>-3</td>
<td>49</td>
<td>6</td>
<td>Left precentral</td>
</tr>
</tbody>
</table>

Note. Voxel level threshold $T = 2.57$, $P < 0.001$ uncorrected and 0.05 corrected for multiple comparisons, spatial extent threshold $>80$ voxels. Bold numbers correspond to a global peak of the cluster, italic—to local peaks within same contiguous cluster.
Differential activation during the Lie also included the aspect of the right SFG (BA 8) contiguous with the ACC, suggesting their functional continuity during the GKT deception (Koski and Paus, 2000). Primate studies have demonstrated rich projections between the BA 8 and the ACC as well as the inhibitory role of BA 8 in previously learned forelimb movements (Oishi and Kubota, 1990; Bates and Goldman-Rakic, 1993). Increased activity at the junction of the left dorsal premotor and prefrontal cortices and the anterior parietal cortex may be related to increased demand for motor control directing right thumb to the appropriate response button during the Lie button press. We speculate that this increase in activation reflects additional effort needed to “overcome” the inhibited true response. This hypothesis could be tested in future studies by systematically varying side, and perhaps modality, of response. While we found brain regions that were more active during Lie than Truth, there were no regions more active during Truth than Lie, suggesting that Truth is the baseline cognitive state.

Our version of the GKT was intended to minimize anxiety response, while maintaining the motivation to deceive with modest positive reinforcement. We did not find activation of the regions frequently associated with positive skin conductance response, anxiety, or emotion (orbitofrontal cortex, lingual and fusiform gyrus, cerebellum, insula, and amygdala) and our participants had no subjective anxiety during the GKT (Gur et al., 1987; Chua et al., 1999; Critchley et al., 2000). Thus, ACC activation during deception is probably not a correlate of anxiety. However, because parts of the ACC may be involved in emotional information processing, our data do not definitively exclude anxiety or emotion-related activation (Whalen et al., 1998). SFG activation has been associated with orientation to the contralateral field induced by cognitive activity (Gur et al., 1983) and its occurrence on the right is consistent with the association of right anterior activity with avoidance and negative affect (Davidson et al., 1990).

The study has a number of limitations stemming from paradigm design and the constraints imposed by the MRI environment. First, under “field” conditions, deception involves elements of choice and more elements of risk and emotion than is the case in our task. Supplementing the GKT with a paradigm that allows the participant a choice in manipulating risk could reveal additional regions of deception-specific activation, such as the orbitofrontal cortex (Bechara et al., 2000). Because a susceptibility artifact limits BOLD fMRI imaging of the orbitofrontal cortex, alternative imaging sequences may be necessary. Second, the 12-s intertrial interval of the event-related design limited the number of stimuli in a single session and thus the statistical power. The repetition of the Lie and Truth stimuli was necessary to amplify the inherently low power of event-related BOLD fMRI paradigms (Aguirre, 1999). Using polygraph, Elaad reported no decline in the accuracy of detection of deception with repetitive GKT stimuli (Elaad and Ben-Shakhar, 1997). Our GKT was controlled for both habituation and the “oddball” effect by equal repetition of all stimuli included in the analysis (Control, Lie, Truth). A modified event-related paradigm with faster stimuli presentation rate and variable intertrial interval (“jitter”) could allow a reduction in repetition of salient stimuli (Burock et al., 1998). Third, the Truth and Lie cards (Fig. 1) differed in both suit and number. Shape and color discrimination have been associated with parietal and occipital but not cingulate activation, making the graphic differences between the Truth and the Lie cards unlikely causes of ACC activation (Farah and Aguirre, 1999). Replication of our findings with a GKT using playing cards that differ in number only or simple numbered cards could resolve this question. Finally, our MRI data have not been correlated with ERP or polygraph recordings. Simultaneous ERP and MRI recording is hampered by the strong magnetic field and is a focus of current research (Goldman et al., 2000). We did not use polygraphy during MRI because of its limited reliability (Office of Technology Assessment, 1990).

CONCLUSION

This report demonstrates a within-group difference between lying and telling the truth using event-related fMRI and the GKT model of deception. This finding indicates that there is a neurophysiological difference between deception and truth at the brain activation level that can be detected with fMRI. The anatomical distribution of deception-related activation indicates that deception involves conflict with, and suppression of, the prepotent (truthful) response. Further refinements of the paradigm design and image analysis methodology could increase the salience and the statistical power of the simulated deception paradigms and establish an activation pattern predictive of deception on an individual level.

ACKNOWLEDGMENTS

Dr. Ehrman helped design the GKT and reviewed the manuscript. Dr. Maria Langleben provided the classical definitions of deception.

REFERENCES


A small portion of the text is as follows:


The full text of the document can be found in the original PDF.