

F. Andrew Kozel,^{1,2,3} M.D., M.S.C.R.; Kevin A. Johnson,^{2,3} Ph.D.; Emily L. Grenesko,^{2,3} B.A.; Steven J. Laken,⁴ Ph.D.; Samet Kose,^{2,3} M.D.; Xinghua Lu,⁵ M.D., Ph.D.; Dean Pollina,⁶ Ph.D.; Andrew Ryan,⁶ Ph.D.; and Mark S. George,^{2,3} M.D.

Functional MRI Detection of Deception After Committing a Mock Sabotage Crime*

ABSTRACT: Using Blood Oxygen Level Dependent (BOLD) functional MRI (fMRI) to detect deception is feasible in simple laboratory paradigms. A mock sabotage scenario was used to test whether this technology would also be effective in a scenario closer to a real-world situation. Healthy, nonmedicated adults were recruited from the community, screened, and randomized to either a Mock-crime group or a No-crime group. The Mock-crime group damaged and stole compact discs (CDs), which contained incriminating video footage, while the No-crime group did not perform a task. The Mock-crime group also picked up an envelope from a researcher, while the No-crime group did not perform this task. Both groups were instructed to report that they picked up an envelope, but did not sabotage any video evidence. Participants later went to the imaging center and were scanned while being asked questions regarding the mock crime. Participants also performed a simple laboratory based fMRI deception testing (Ring-Watch testing). The Ring-Watch testing consisted of "stealing" either a watch or a ring. The participants were instructed to report that they stole neither object. We correctly identified deception during the Ring-Watch testing in 25 of 36 participants (Validated Group). In this Validated Group for whom a determination was made, computer-based scoring correctly identified nine of nine Mock-crime participants (100% sensitivity) and five of 15 No-crime participants (33% specificity). BOLD fMRI presently can be used to detect deception concerning past events with high sensitivity, but low specificity.

KEYWORDS: forensic science, lie detection, functional MRI, deception, mock crime

Intense interest exists in the scientific community and lay press concerning the possibility of using Blood Oxygen Level Dependent (BOLD) functional magnetic resonance imaging (fMRI) to measure brain activation during deception. A number of studies using fMRI to investigate the neural correlates of deception have been published (1–14). The design and analysis methods across these studies vary considerably, making it difficult to integrate the results. At the group analysis level, however, these studies have consistently found significant brain activation in deception versus telling the truth. There has been variability in the specific brain regions activated during deception in these studies. One explanation for this array of findings is the diversity in tasks and questioning paradigms. To date, successful individual analysis has only been achieved in two studies (2,5). The University of Pennsylvania group (Langleben et al.) also reported on using a different analytic approach to the same imaging data to improve their accuracy (12).

¹Department of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9119.

²Center for Advanced Imaging Research (CAIR), Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425.

³Brain Stimulation Laboratory, Department of Psychiatry, Medical University of South Carolina, 67 President Street, P.O. Box 250861, Charleston, SC 29425.

⁴Cephos Corp., P.O. Box 45, Tyngsboro, MA 01879.

⁵Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina, 135 Cannon Street, Suite 303, Charleston, SC 29425.

⁶Defense Academy for Credibility Assessment (DACA) [formerly Department of Defense Polygraph Institute (DODPI)], 7540 Pickens Avenue, Fort Jackson, SC 29207.

*This work was presented as a poster at the 13th Annual Meeting of the Organization for Human Brain Mapping in Chicago, June 2007.

Received 28 Oct. 2007; and in revised form 16 Mar. 2008; accepted 17 Aug. 2008.

Although these studies reported reasonably high individual accuracy rates, there are several concerns which must be addressed prior to moving this technology to real-world application. One concern is that relatively simple deception paradigms (theft of a watch or ring, deception about which playing card one is holding) were used with perceived financial compensation for successful deception. The impact of testing for deception on more elaborate scenarios that are closer to a real-world situation is unknown. Importantly, the robustness of our methodology using *a priori* defined regions of interest (Kozel et al., 2005) is untested for detecting deception when performing different tasks and providing different types of lies. Another concern is the time between the act in question and testing (3). The ring/watch and playing card studies used a laboratory-based scenario performed immediately before the MR scanning to test for deception. To determine how robust our prior published method (5) was in detecting deception in a more real-life setup with extended time from the event, we worked with the Defense Academy for Credibility Assessment (DACA) to develop a mock sabotage (i.e., the destruction of property or obstruction of normal operations) crime paradigm. This paradigm was closer to a real-world situation in both the nature of the criminal act and the time between the act and testing. We hypothesized that we would be able to determine which participants committed the crime versus those who did not commit the crime using our automated computerized analysis methods on the fMRI data to detect deception. In addition, we hypothesized that using a laboratory screening deception test would improve our predictive power for the mock crime. We reasoned that we would have a higher predictive power in individuals for whom we were able to properly identify lies versus truth on an embedded laboratory study of deception (i.e., Ring-Watch task). That is, using the lab-based identification as a within subject assay might improve our ability to

detect deception in the more elaborate Mock-crime task if we only made Mock-crime calls in the subjects that we correctly identified in the simple laboratory study.

Methods

Visit 1

Participants were healthy, literate adults between the ages of 18–50 years who were recruited by written advertisements from within the Medical University of South Carolina (MUSC) and the Charleston, SC community. (See Fig. 1 for outline of study flow.) Informed consent approved by the MUSC Office of Research Integrity was obtained in writing. Participants were screened with a Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (15), a pre-MRI screening form, a medical history, and were given a brief physical exam. Exclusion criteria included taking any medications within five half-lives of imaging, presently using illegal

drugs, history of a psychiatric disorder except simple phobia, history of a significant central nervous system disease (e.g., stroke, seizures, severe head injury, etc.), currently unstable medical condition, pregnancy or lactating, caffeinism (i.e., headaches or other withdrawal symptoms with cessation of caffeine for 3 days), nicotine use, claustrophobia, previous inability to tolerate an MRI, any metal implants making imaging unsafe, or prior knowledge of the paradigm. All participants were also evaluated with an Annette Handedness Scale (16), the State-Trait Anxiety Inventory (STAI) (17), and the Temperament and Character Inventory (18). Prior to MRI scanning a urine sample was obtained to test for drugs of abuse as well as pregnancy testing for women of childbearing potential. Participants received \$20 for completing the screening phase. For those participants that met the screening criteria, they were randomly assigned to either the Mock-crime group or No-crime group. Participants were then scheduled for a second visit in which they were given instructions and performed the assigned tasks.

Subject randomization was performed at DACA by Dean Pollina (DP) using an urn randomization controlling for sex and age (<35 or ≥35). The group assignment was placed in sequential envelopes, sealed, and sent to Emily Grenesko (ELG) at MUSC. The entire study team, except ELG, was unaware of the randomization scheme used or the percentage committing the mock crime (i.e., base rate). Once participants were screened and enrolled, ELG would open the next envelope in the appropriate category (e.g., female and >35 years) to determine group assignment.

Visit 2

The mock-crime scenario used in the present study was adapted from procedures that were used to develop computerized evaluations of polygraph data (19). Based on internal studies performed at DACA and the constraints of performing an fMRI experiment, a number of modifications to the paradigm were made. Very similar paradigms have been used by DACA to test other modalities designed to detect deception (personal communication A. Ryan).

Mock-Crime Group—Participants received instructions from ELG regarding how to carry out the mock crime. They were told to first pick up an envelope marked “Confidential” from a research scientist in room 370 in the Main Hospital (different building from both the location of screening and where the mock crime was performed). They were instructed not to open it and bring it back to ELG before their MRI scan. The subject was then told that their next task would be to destroy evidence of a crime. The explanation was that a surveillance camera captured someone else committing a crime and that person wanted the evidence destroyed to avoid prosecution. In order to do this, they needed to go to room 112 in the Forensic Psychiatry Building where the evidence was being kept for processing. They were instructed to spend as little time in the room as possible—30 min at the most. They were told to use a specific entrance to the building across from the law enforcement parking spaces. Once in the room, they were to find a compact disk (CD) in a specific folder, view the CD to ensure that they saw a robbery of a convenience store taking place, and then destroy the CD by any means possible before leaving the room. A shredder in the room was suggested for use, but it purposely did not work. Thus, they were required to deal with an unexpected situation and determine how to break the CD using another method. The purpose was to add some unpredictability to the scenario and increase the sense of jeopardy and stress. Participants were to retain a small piece of the CD and bring it to ELG just before their MRI scan.

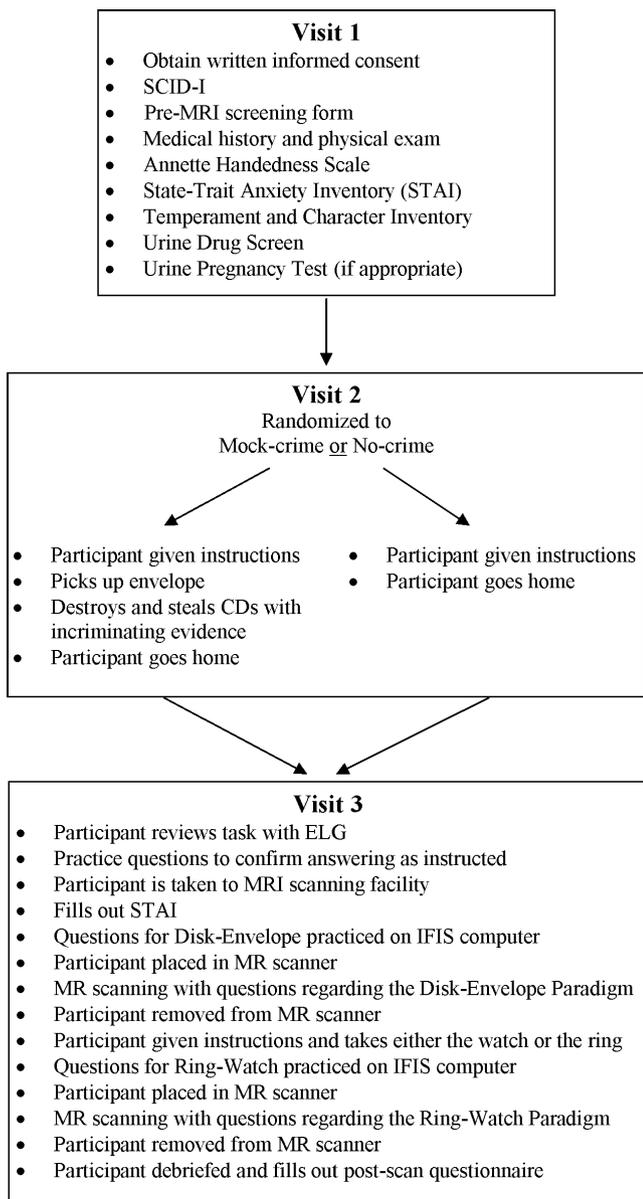


FIG. 1—Diagram of participant flow.

The subjects were then told by ELG that she needed them to do something in addition “just for her” while they were in that room. She said that she had taken a laptop from one of her professors some time ago but had recently noticed security cameras in his office. She told them a CD containing this incident was in another separate folder in the room. They were instructed to view the CD to ensure that they saw ELG committing the crime and to bring it back to her intact before their scan. The purpose of this additional task was to add more of a sense that the subject was actually committing a crime. The room containing the CDs had hidden video surveillance monitoring equipment and a webcam to confirm and document that the participants accurately performed the task. The room also had a telephone that ELG called while the participants were performing the task in the room. The phone call was intended to add to the sense of possibly getting caught in the act. Participants were informed that they would be contacted by ELG to set up Visit 3 if the task was performed properly. Participants received \$50 compensation for completing the task portion of the study. If the task was not performed properly, the subject was compensated for Visits 1 and 2 and excluded from the study. Throughout the remainder of the study, the participants were to report that they went to pick up the envelope but did nothing else (i.e., they were to lie about the Mock-crime task). Any deviation from that would result in the subject being excluded from the study and compensated for Visit 1 and 2.

No-Crime Group—ELG ensured that the participants understood that someone destroyed evidence on a CD of a crime being committed. They were told they would need an alibi as to where they were during that time. For their alibi, they were instructed to say that they were in room 370 in the Main Hospital picking up an envelope marked “Confidential” from a research scientist. The participants were informed that they would not actually do this task. Throughout the remainder of the study, they were to answer as if they picked up the envelope but did not damage the CD or visit the crime room. The purpose of having the participants lie about picking up the envelope was to have questions during the fMRI evaluation in which the participants in this group were lying. Participants were allowed to leave after the instructions and were subsequently contacted for their MRI scan (i.e., Visit 3). They were instructed that they would receive \$50 for performing Visit 2.

Visit 3

Review of Task—Participants met ELG prior to going to the scanner. If the participant had been assigned to the Mock-crime group, they were expected to produce the envelope, a piece of the CD which contained video of the convenience store robbery, and the intact CD which contained video of ELG stealing the laptop. ELG confirmed that all participants were abiding by the story of picking up the envelope but not damaging the CD. The questions asked in the scanner were reviewed to confirm that the subject was answering per protocol.

Mock-Crime fMRI Testing—The subject was escorted to the scanner by ELG who then left for the remainder of the scanning. They were given the STAI to fill out and were trained by Kevin Johnson (KAJ) on answering the questions using the IFIS system (Intermagetics General Corp., Latham, NY) outside of the scanner. Subsequently, the participants were made safe to enter the scanning room and were positioned in the MRI scanner with the IFIS system signaling gloves. In addition, physiologic measuring devices were attached to collect data on pulse, breathing rate, and electrodermal

response (EDA, galvanic skin response) every 0.01 sec during their examination (data not included). The participants were given foam ear protection and had their head secured using foam pads. All scanning was performed using a Philips 3.0 Tesla Intera MRI scanner equipped with an actively shielded magnet. The BOLD fMRI Echo Planar Imaging (TR 1867 ms, TE 30 ms, Flip Angle 90 deg, FOV 208 mm, matrix 64 × 64, SENSE factor 2, 36 slices, 3 mm with 0 mm gap, giving a voxel size of 3.25 × 3.25 × 3.00 mm³) covered the entire brain and was positioned with reference to the Anterior Commissure-Posterior Commissure line using a sagittal scout image. A quick motor task (3 min—97 volumes) was performed to orient participants to the fMRI environment and to give them practice responding with the IFIS glove (yes—right thumb, no—right index finger). For the Mock-crime testing, BOLD fMRI scans were acquired (12 min—386 volumes) while questions were visually displayed. Using an event-related design, three types of questions were asked regarding whether the subject picked up the envelope, whether the subject committed the sabotage, or general neutral questions. Questions were presented visually for 3.5 sec. After a question, there was a visual prompt for the subject to answer “Yes or No” for 2 sec followed by a “+” for 0.5 sec. Participants were instructed to not answer until they saw the visual prompt of “Yes or No.” The delayed response was to reduce the variability of response timing due to differences in reading speeds across questions and across participants. Because of this forced delay to respond, the reaction times cannot be interpreted as indicating the time it required to answer the question. Thus, each question with answer epoch took 6 sec. There were 20 questions for each category (60 unique questions total): “disk” regarding the Mock-crime; “envelope” regarding picking up the envelope; and “neutral” regarding easily answered and verifiable personal information. The order of the type of question (i.e., disk, envelope, and neutral) was pseudorandomized using a web-based randomization generator (<http://www.randomization.com>) but consistent across participants. The IFIS system pseudorandomly chose the actual question for the appropriate category. Thus, the order of the type of questions was consistent across participants, but the order of the actual questions asked was varied. The 60 questions were presented once for practice outside of the scanner and in two separate sets (total 120 questions—each question with answer epoch lasting 6 sec) that were administered consecutively for the scanning. Participants were then taken out of the scanner to participate in the task of taking the ring or watch.

Ring-Watch fMRI Testing—Upon completion of the Mock-crime paradigm questions, participants were given a brief break before performing the Ring-Watch paradigm (similar to paradigm in prior study [5]). For the Ring-Watch paradigm, the subject was instructed to go into a room where there were two objects hidden in a drawer. The instructions were to steal one of the objects and leave the other behind. A research assistant confirmed that only one object was taken. The participant was instructed to respond to questions in the scanner as if he/she had not taken either object. Once in the scanner, BOLD fMRI scans (same scanning parameters as in the Mock-crime task, 386 volumes) were acquired while questions were visually displayed in the same manner as the Mock-crime testing. The number of questions presented during the Ring-Watch testing was equal to the number of questions presented for the Mock-crime testing. The difference was that the questions consisted of different neutral questions, and whether the subject took the ring or took the watch. All questions were pseudorandomized and presented twice. The IFIS gloves were used to respond “yes” or “no” to the questions. Unlike in the Kozel et al.

2005 study, the current study participants were not encouraged or given monetary incentives to “beat” the test and control questions were eliminated. A structural scan was subsequently acquired and reviewed for any gross pathology. When the task was completed, participants were removed from the scanner and debriefed. Scanning time in total was *c.* 40 min, but the entire task took about 1.5–2.0 h for completion. KAJ documented which object the subject took. KAJ and ELG were the only investigators in the study who knew which object was taken.

Debriefing Process—After participants were removed from the MRI, they were greeted by ELG who asked if they had any questions about the protocol or what just occurred. They were asked if they experienced any difficulties during the study or scanning. Participants were given a postscanning questionnaire and paid \$50 for completing the scanning. One of the questions in the postscanning questionnaire specifically asked the participants whether they felt that the sabotage mock-crime scenario was believable. Once again, only ELG knew which task the subject actually performed; and only KAJ and ELG knew whether the watch or ring was taken for each subject.

Data Analysis

All initial imaging data analysis was performed blind to subject group. After initial data analysis, datasets were locked, and the blind broken. There was no possibility of any data analysis by anyone with knowledge of the group assignment. For the Ring-Watch task, KAJ revealed the blind to SJL, MSG, and FAK. Subsequently, the assignment of Mock-crime versus No-Crime task was unblinded at a later date with all of the above parties with the addition of DP.

Participants

Testing of demographic and clinical variables was performed using *t*-test for continuous data and Chi-square for discrete data to determine if there were any significant differences between the two groups.

Behavioral Data

Participant responses were inspected for participation and irregularities with respect to the correct answer per the protocol. Response data from the IFIS was converted to Excel files using E-Prime 1.1.4.1 which were subsequently imported into Matlab (The Mathworks, Natick, MA). Using a Matlab script, response data from the IFIS files were converted into the onsets for the event-related SPM2 analysis. Responses that were not consistent for both times the question was asked, not answered, or not answered as specified in the protocol were identified and modeled as separate “nonprotocol” events. Although the reaction times do not indicate the time required answering different question types, differences in reaction times between the Mock-crime group and No-crime group could indicate differences in the group’s testing performance. Response times as measured by the IFIS were compared using a two-sided *t*-test for the two groups using NCSS (20).

Functional MRI Analysis

Functional MRI analysis was carried out using Statistical Parametric Mapping software (SPM 2; Wellcome Department of Cognitive Neurology, London, UK) by Steven J. Laken (SJL) and F. Andrew Kozel (FAK) independently. The two analyses differed

in the following ways: FAK performed SPM2 analysis using Red Hat Linux Enterprise Edition (Linux kernel 2.6) and Matlab version 7.1.0.183 (Release 14 with service pack 3) and SJL performed SPM2 analysis using Windows XP x64 and Matlab version 2006a. All analyses, Mock-crime and Ring-Watch were carried out with SJL and FAK blind to the actual tasks performed by the participants. Data analysis was carried out in the same manner as in Kozel et al. 2005 except statistical estimation was modeled to correct for temporal coherence using AR(1) (21) and the hemodynamic response function included a temporal derivative. We felt that the use of the AR(1) decreased the impact of temporal dependence, and the use of a temporal derivative enabled a broader hemodynamic response to be fitted to our model. Further all data from Kozel et al. 2005 was tested using these changes and did not change the results (data not shown). The first image was reoriented using the display function in SPM2 so that the 0,0,0 coordinate corresponded to the medial anterior commissure. Each scan was then adjusted to set these coordinates using the reorient function in SPM2. SJL and FAK set these points independently. Preprocessing was performed using an automated script as described in Kozel et al. 2005. The reoriented images were realigned and unwarped to correct head movement and resulting EPI distortions. Participants with movement of >3 mm were eliminated. Slice timing was performed to correct for differences in the point in time that each slice was acquired. The functional images were then spatially normalized to the SPM EPI template and re-sampled to $3 \times 3 \times 3$ mm voxels (22). The data were spatially smoothed using a Gaussian kernel with 8 mm full width at half maximum based on the suggested standard of 2 to 3 times the output spatially normalized voxel size. This was performed to adjust for inter-subject variability as well as making the errors more normal in their distribution to help ensure the validity of inferences based on parametric tests (23).

Using a Matlab script, the statistical portion of the analysis was performed. A general linear model within SPM2 was specified and estimated for the Mock-crime and the Ring-Watch paradigms. Events and their temporal derivative were defined as occurring when the cue to answer “Yes” or “No” was presented to the participants (starting at 3.5 sec and occurring every 6 sec thereafter). Effects at each and every voxel were estimated using the general linear model at the first statistical level. The motion-recorded parameters generated during the “Realign” process were included as six user-specified regressors. The nonprotocol events were also included as conditions and modeled with the hemodynamic response function. A high pass filter (cut-off frequency = 128 s) was used to remove possible effects of low-frequency changes and AR(1) was used to decrease the impact of temporal dependence. Individual *t*-statistics activation maps were defined based on the contrasts of interest. For the Mock-crime analysis, the contrasts were disk minus neutral and envelope minus neutral. For the Ring-Watch analysis, the contrasts were ring minus neutral and watch minus neutral. Using a Matlab script, the number of significant ($p < 0.05$) voxels was determined in each of the three Regions of Interest (ROI) defined in Kozel et al. 2005 (clusters 1, 2, and 4 which roughly correspond to right anterior cingulate region, right orbitofrontal/inferior frontal region, and right middle frontal region). For the *Mock-crime Analysis*—the number of significantly activated voxels for the disk minus neutral contrast was subtracted from the envelope minus neutral contrast using ROIs 1, 2, and 4. If the resulting value was positive, then the call was made that the mock crime was committed (i.e., greater brain activation corresponding to lying about performing the mock crime). If the resulting value was zero, then it was called indeterminate. If the

resulting value was negative, then the call was that the mock crime was not committed (i.e., greater brain activation corresponding to lying about the envelope task). Similarly for the *Ring-Watch Analysis*—the number of significantly activated voxels for the ring minus neutral contrast was subtracted from the watch minus neutral contrast using ROIs 1, 2, and 4. If the resulting value was positive, then the call was made that the ring was taken (i.e., greater brain activation corresponding to lying about the ring questions). If the resulting value was zero, then it was called indeterminate. If the resulting value was negative, then the call was that the watch was taken (i.e., greater brain activation corresponding to lying about the watch questions).

Imaging Data Quality Check—The functional imaging data was checked for gross artifacts that would preclude analysis and motion >3 mm in any plane. The behavioral data was checked to ensure that <10 responses per group (e.g., disk) were “not per protocol” (i.e., incorrectly answered or not answered). Also, participants that did not adhere strictly to the protocol were eliminated. Once FAK and SJL had completed their blinded analyses, the results were compared. If each reached a different conclusion, then that subject was analyzed again from the beginning by both investigators. If the conclusions drawn were still different, then that subject was eliminated. The participants that were *not eliminated* due to excessive motion, gross artifacts, inadequate number of correct responses, not performing the protocol correctly, or lack of agreement in analysis are referred to as the “Quality Group.”

The *a priori* defined primary analysis of correctly identifying deception in the Mock-crime group was based on the subset of the Quality Group for which the Ring-Watch analysis correctly identified the object taken. This group is referred to as the “Validated Group” since the technology was validated to work on a known condition for this individual during this scanning session (i.e., the lie and truth are known depending on whether the ring or watch was taken). Finally, all participants were included into secondary analyses using the Quality Group and then the entire group without restrictions to determine which, if any, of these factors influenced the ability to detect mock-crime deception.

ROC Analysis—Receiver operator characteristic curves were generated using NCSS (20) by using the number of voxels with significantly activated *t*-values greater than or equal to 1.645 from the contrasts ([Disk – Neutral] – [Envelope – Neutral]) for the clusters 1, 2, and 4 as described in Kozel et al. 2005. Participants were segregated into Mock-crime groups for sensitivity values and No-crime groups for specificity values. Sensitivity and specificity was calculated for each group by stepping through each of these values for the Complete Group, Quality Group, and Validated Group.

Group Imaging Data Analysis—A number of group analyses were performed to increase our understanding of the relationship of this study to prior work. Using the individual *Lie-minus-True* and *True-minus-Lie* contrast images produced at the first statistical level analysis for the Mock-crime paradigm and Ring-Watch paradigm, group *t*-maps were generated at the second level using a random effects model (24). Only participants who met the criteria for the Quality Group were included in the analysis. Significance was defined as False Discovery Rate > 0.05 with a cluster size of ≥25 voxels which was the threshold level used in Kozel et al. 2005. For the Mock-crime paradigm, group analyses were performed for the entire Quality Group as a whole, those in the subset of the

Mock-crime group, and those in the subset of the No-crime group. For the Ring-Watch paradigm, group analyses were performed on the entire Quality Group.

Results

Participants

Seventy participants were enrolled of which 13 were eliminated at screening (see Fig. 2). This left 57 participants who were randomized to either the Mock-crime group ($n = 27$) or No-crime group ($n = 30$). After randomization, there were five participants in the Mock-crime group and four participants in the No-crime group who dropped out. The remaining 48 ($n = 22$ Mock-crime, $n = 26$ No-crime) participants that were scanned will be referred to as the “Complete Group.”

To address the question of how believable was the study paradigm for the participants, the postscanning questionnaire asked the subjects, “Did you feel this scenario was believable?” For the Complete Group scanned, 43 of 48 (90%) participants answered, “Yes.” Thirty-two of thirty-six (89%) in the Quality Group and 22 of 25 (88%) in the Validated Group answered, “Yes.” Thus, the participants generally felt that the scenario was believable.

The average time for the Mock-crime group from performing the mock crime (visit 2) to fMRI scanning was 105.0 h (90.6 SD, range 5.5–312.0) in the Validated Group. The average time for the No-crime group from getting instructions (visit 2) to fMRI scanning

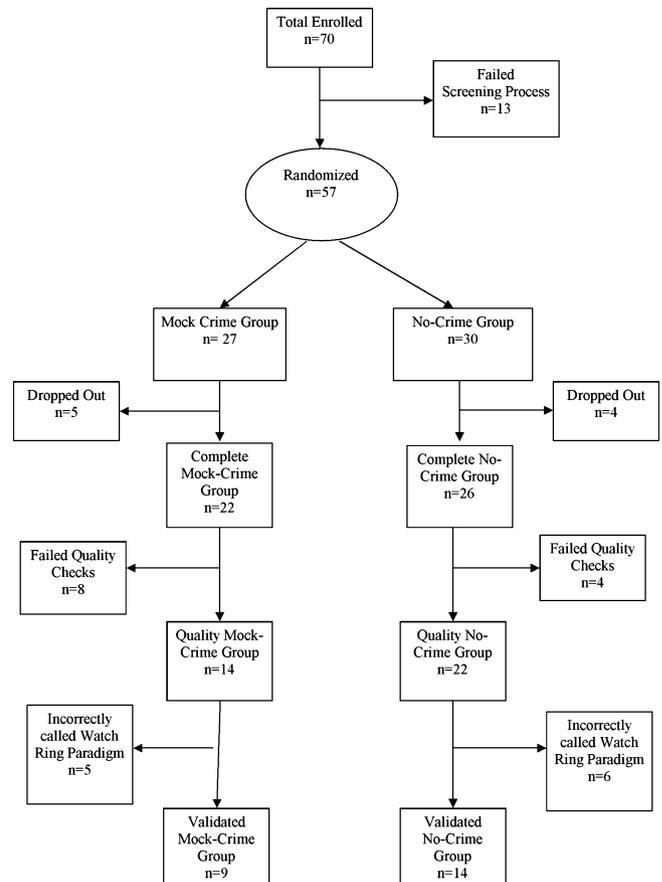


FIG. 2—Schematic of the number of subjects enrolled, randomized, and scanned. The breakdown of the numbers of participants for the Complete Group, the Quality Group, and the Validated Group are indicated for the subjects that were randomized to the Mock-crime and No-crime tasks.

TABLE 1—Demographics of participants.

	All Subjects			Envelope			Disk			Chi-Square or <i>t</i> -Test			
	Complete	Quality	Validated	Complete	Quality	Validated	Complete	Quality	Validated	C–C	Q–Q	V–V	Statistics
Ethnicity													
C	35	27	20	22	18	14	13	9	6	6.02	2.86	0.60	Chi-square
A	1	1	0	1	1	0	0	0	0	0.05	0.24	0.74	<i>p</i> -value
AA	12	8	5	3	3	2	9	5	3				
Object taken													
Took ring	26	21	16	14	11	10	12	10	6	0.00	1.62	1.21	Chi-square
Took watch	22	15	9	12	11	6	10	4	3	0.96	0.20	0.27	<i>p</i> -value
Ages													
Avg. age	30.8	30.8	31.6	31.2	31.3	32.6	30.2	29.9	29.8	−0.34	−0.40	−0.65	<i>t</i> -value
SD	9.6	10.0	10.4	10.0	10.8	11.4	9.3	8.9	8.6	0.73	0.69	0.52	<i>p</i> -value
Low	18	18	18	18	18	18	19	19	19				
High	50	50	50	50	50	50	50	44	43				
Gender													
M	25	18	12	13	10	7	12	8	5	0.10	0.47	0.32	Chi-square
F	23	18	13	13	12	9	10	6	4	0.75	0.49	0.57	<i>p</i> -value
Employment													
FT	5	3	2	1	0	0	4	3	2	4.15	6.14	2.13	Chi-square
FT-MUSC	19	14	10	11	9	7	8	5	3	0.39	0.19	0.55	<i>p</i> -value
PT	1	1	1	1	1	1	0	0	0				
PT-MUSC	1	1	1	1	1	1	0	0	0				
Student	22	17	11	12	11	7	10	6	4				
Handedness													
L	3	3	3	2	2	2	1	1	1	0.20	0.04	0.18	Chi-square
R	45	33	22	24	20	14	21	13	8	0.65	0.84	0.67	<i>p</i> -value
Education													
Avg. years	16.0	15.9	15.7	15.9	16.0	15.7	16.2	15.7	15.9	0.51	−0.42	0.29	<i>t</i> -value
SD	2.0	2.0	1.9	2.2	2.1	1.9	1.8	1.8	2.0	0.62	0.68	0.78	<i>p</i> -value
Min	12	12	12	13.0	13.0	13.0	12	12	12				
Max	20.5	20.5	18	20.5	20.5	18.0	20	18	18				
Time to scan													
Avg. time	78.9	86.9	68.5	56.8	60.4	51.5	105.0	128.6	98.7	2.12	2.44	1.60	<i>t</i> -value
SD	81.3	87.2	72.9	66.4	66.6	57.6	90.6	101.2	90.2	0.04	0.02	0.12	<i>p</i> -value
Min	0.5	0.5	0.5	0.5	0.5	0.5	5.5	24.0	24.0				
Max	312.0	312.0	312.0	216.0	216.0	216.0	312.0	312.0	312.0				

Under ethnicity, C, A, and AA represents Caucasian, Asian, and African-American. Object taken reflects the number of subjects who chose to take a ring or a watch. M and F represents male and female. FT, FT-MUSC, PT, PT-MUSC, and student represents full-time worker, full-time at MUSC, part-time, part-time at MUSC, and full-time student, respectively. L and R represent left and right. Education is the number of years the person has spent in school. Time to scan is from the time they met with ELG for Visit 2 to the time of their fMRI scan. Statistical comparisons were made between the Complete No-crime versus Complete Mock-crime (C–C), Quality No-crime versus Quality Mock-crime (Q–Q), and Validation No-crime versus Validation Mock-crime (V–V) groups using either a Chi-square test or a two-sample two-way *t*-test. Statistically significant *t*-values and Chi-square values are shown highlighted in bold. The only significant differences were for ethnicity in the C–C comparison and time to scan for the C–C and Q–Q comparisons.

was 56.8 h (66.4 SD, range 0.5–216.0) in the Validated Group (see Table 1). The difference in time from performing the task (or given instructions of what occurred) and the scan was significantly different between the Complete Mock-crime and Complete No-crime groups (two sample *t*-test, two-sided, $t = 2.12$, p -value = 0.04) and between the Quality Mock-crime and Quality No-crime groups (two sample *t*-test, two-sided, $t = 2.44$, p -value = 0.02). There were, however, no significant time differences between the Mock-crime and No-crime Validation Groups (two sample *t*-test, two sided, $t = 1.60$, p -value = 0.12). Although there was a difference in time to scan for the two groups, there was no significant difference (two sample *t*-test, two-sided, $t = 0.52$, p -value = 0.61) in time to scan for those participants accurately and inaccurately called.

No significant differences were found in comparisons of ethnicity, object chosen to steal (i.e., ring or watch), gender, handedness, and years of education between the Mock-crime and No-crime groups for the Complete Group, the Quality Group and the Validation Group (see Table 1). There was, however, a difference between the two groups with respect to reaction times for all question types. The No-crime group responded significantly faster to all question types than the Mock-crime group ($p < 0.001$ for Complete, Quality, and Validated groups).

Functional MRI Individual Lie Detection Testing

Of the 48 participants in the Complete Group, twelve ($n = 8$ Mock-crime group, $n = 4$ No-crime group) did not meet criteria to be in the Quality Group. The reasons for exclusion from the Quality Group were varied. Three participants in the Mock-crime group did not complete the protocol properly and one of these also had inadequate number of per protocol responses recorded. Three participants had motion >3 mm on the Ring-Watch testing with one of these also having >3 mm motion on the Mock-crime testing. Three participants had significant artifacts on their EPI scans for both the Mock-crime and Ring-Watch testing. In addition to the artifacts, one of these participants had an inadequate number of per protocol responses recorded. Another subject had an inadequate number of per protocol responses recorded. There were two participants in which SJL and FAK did not agree on their calls. Both participants had 0 or 1 voxels activated as determined by SJL or FAK which most likely represented spurious results. The Quality Group was comprised of 14 participants in the Mock-crime group and 22 participants in the No-crime group.

Mock-Crime Paradigm Testing—To determine the Validated Group, those participants in the Quality Group who were not

correctly identified or indeterminate on the Ring-Watch testing were eliminated from the Validated Group ($n = 5$ for the Mock-crime group and $n = 6$ for the No-crime group). For the Validated Group, nine of the nine assigned to the Mock-crime were correctly identified, and five of the 16 who were in the No-crime group were correctly identified with one of the calls being an indeterminate. For those in the Validated Group for whom a call was made, this resulted in a sensitivity of 100% (9/9 correct classification, 95% confidence limit = 0.68–1.00) and a specificity of 33% (5/15 correct classification, 95% confidence limit = 0.15–0.58) for detecting the mock crime (see Table 2).

For the Quality Group, thirteen of the fourteen Mock-crime participants were correctly identified and eight of the 22 No-crime participants were correctly identified (1 was an indeterminate call). For those in the Quality Group for whom a call was made, this results in a sensitivity of 93% (13/14, 95% confidence limit = 0.69–0.99) and a specificity of 38% (8/21, 95% confidence limit = 0.20–0.59). Thus, restricting judgments to only those identified correctly in the lab study did not significantly ($p > 0.4$) improve the overall detection rates.

For the Complete Group, 20 of the 22 Mock-crime group participants were correctly classified. The No-crime group was correctly identified 10 out of 26 times with two of those calls being indeterminate. For those in the Complete Group for whom a call was made, the sensitivity was 91% (20/22, 95% confidence limit = 0.72–0.97) and the specificity was 42% (10/24, 95% confidence limit = 0.22–0.61).

Ring-Watch Paradigm Testing—For the Quality Group, the Ring-Watch paradigm correctly identified the deception regarding which item was taken 25 out of 36 times with one subject having an indeterminate call. This results in an accuracy of 71% (25/35, 95% confidence limit = 0.55–0.84). For the Complete Group, Ring-Watch paradigm deception was correctly identified 72% (31/43, 95% confidence limit = 0.57–0.83) of the time with 5 individuals having indeterminate results.

ROC Curves—We used receiver operator characteristic (ROC) curves to systematically test the trade-off between sensitivity and specificity. For our *a priori* analysis, we used a threshold of the significant ($t > 1.645$) voxel number difference greater than zero as being the lie. The area under the curve (AUC) is calculated to determine the overall accuracy of the test and the AUC can be compared to a test that is no better than chance. For each group, the BOLD fMRI test discriminated better than chance (see Table 3

TABLE 2—Sensitivity and specificity for mock-crime task.

	Committed Mock Crime	No Mock Crime
Validated group, $n = 25$ (There was one indeterminate in the No Mock Crime Group)*		
FMRI test – Committed crime	9	10
FMRI test – No crime	0	5
Quality group, $n = 36$ (There was one indeterminate in the No Mock Crime Group)†		
FMRI test – Committed crime	13	13
FMRI test – No crime	1	8
Complete group, $n = 48$ (There was two indeterminate in the No Mock Crime Group)‡		
FMRI test – Committed crime	20	14
FMRI test – No crime	2	10

*Sensitivity = 100%, Specificity = 33%.

†Sensitivity = 93%, Specificity = 38%.

‡Sensitivity = 91%, Specificity = 42%.

and Fig. 3). The area under the curve, or accuracies, for the Complete Group was 0.77 (one-sided z -test for AUC = 0.5, z -value = 3.75, $p = 0.0001$), for the Quality Group was 0.77 (one-sided z -test for AUC = 0.5, z -value = 3.3, $p = 0.005$), and for the Validated Group was 0.78 (one-sided z -test for AUC = 0.5, z -value = 2.85, $p = 0.0022$). A Q–Q plot demonstrated that the voxel difference number score values approximated a Gaussian distribution. In addition, a plot of the distribution of the voxel number difference scores demonstrated that the Mock-crime group was largely to the right of zero (i.e., a correct call would be greater than zero) while the No-crime group (i.e., a correct call would be less than zero) was more diffusely distributed (see Fig. 4).

Functional MRI Group Results

Mock-crime Paradigm—The *Lie-minus-True* contrast analysis for the Quality Group failed to find significant (FDR < 0.05, $k > 25$) activation for the entire group. When the groups were split by whether they were in the Mock-crime or No-crime groups, however, both produced areas of significant activation (see Table 4, Fig. 5). The Mock-crime group had a lateral and prefrontal pattern of activation similar to the Model-building group in our prior study, while the No-crime group had a predominantly medial prefrontal activation pattern. A two sample t -test using SPM2 confirmed that there were significantly (FDR < 0.05, $k > 25$) different regions of activation between the Mock-crime and No-crime groups for the *Lie-minus-True* contrast.

The *True-minus-Lie* contrast group analysis of the Quality group for the entire group or split by whether the mock crime was committed failed to find significant activations.

Ring-Watch Paradigm—The *Lie-minus-True* and *True-minus-Lie* contrasts failed to find significant (FDR < 0.05, $k > 25$) activation.

Discussion

Using a unique scenario that more closely approximated a real world situation and performing testing in the MR scanner after a period of delay, our *a priori* defined methodology was found to have good sensitivity (100%) but low specificity (33%). While the overall results are less accurate than in our prior study, we again found that our BOLD fMRI method is significantly better than chance at detecting deception at the individual level. Also, the ROC curves revealed that if the cut-off values are changed (i.e., the value of the difference in the number of activated voxels defined to indicate deception), considerable improvements in specificity can be gained with only modest reductions in sensitivity. Interestingly, in contrast to our prestudy hypothesis, the use of the Ring-Watch Paradigm as an internal study screen to validate whether our technique would work in a particular individual did not drastically improve our diagnostic results in this study.

Our present methodology compares brain activation correlated to responses regarding two conditions (committed mock crime vs. picked up envelope or took the watch vs. took the ring) that are mutually exclusive. One condition (e.g., mock crime) is determined to be the lie because it caused more activation in the previously identified brain regions of interest compared to the other condition (e.g., envelope). This requires the two conditions to be mutually exclusive and the questions to be specific and equally counterbalanced. One explanation for our low specificity is that the cognitive work load for lying about the envelope task was not adequate to obtain a reliable activation. Because of the nature of the testing and

TABLE 3—ROC numbers.

Cut-off Value	Sensitivity (%)	Specificity (%)
<i>Validated group</i>		
-161	100	0
-108	100	6
-99	100	13
-52	100	19
-45	100	25
0	100	31
10	100	38
15	100	44
19	100	50
27	100	56
36	100	63
59	100	69
81	100	75
92	89	75
101	78	75
107	67	75
108	56	75
132	44	75
195	33	75
215	33	81
220	22	81
251	22	88
275	11	88
326	0	88
432	0	94
<i>Quality group</i>		
-108	100	5
-99	100	9
-52	100	14
-45	100	18
-34	100	23
-16	100	32
-1	93	32
0	93	36
2	93	41
4	93	45
10	86	45
15	86	50
19	86	55
27	86	59
36	86	64
43	86	68
59	86	73
81	86	77
92	79	77
101	71	77
103	64	77
107	64	82
108	57	82
132	50	82
150	43	82
187	36	82
195	29	82
215	29	86
220	21	86
251	21	91
275	14	91
326	7	91
336	7	95
432	0	95
<i>Complete group</i>		
-123	100	4
-108	100	8
-102	100	12
-99	100	15
-52	100	19
-45	100	23
-34	100	27
-16	100	35
-6	95	35
-1	91	35

TABLE 3—(Continued)

Cut-off Value	Sensitivity (%)	Specificity (%)
0	91	38
1	91	46
2	86	46
4	86	50
10	82	50
15	82	54
19	82	58
22	82	62
27	77	65
36	77	69
43	77	73
59	77	77
81	77	81
73	73	81
90	68	81
92	64	81
101	59	81
103	55	81
107	55	85
108	45	85
132	41	85
150	36	85
187	32	85
195	27	85
201	27	88
215	23	88
220	18	88
251	18	92
264	14	92
275	9	92
326	5	92
336	5	96
432	0	96

The resulting sensitivity and specificity values are presented for the different cut-off numbers for the Validated, Quality, and Complete groups. The value to determine the call of whether the person committed the mock crime (i.e., the decision number) represents the difference in the number of activated voxels for the Mock-crime minus neutral contrast subtracted by the number of activated voxels for the envelope minus neutral contrast. If the decision number is less than the cut-off value, then the call of no sabotage is made. Conversely, if the decision number is greater than the cut-off value, then a call of committing the sabotage is made. The value of “0” was used in the *a priori* hypothesis in this study and is shown in bold. As can be seen from the above table, small changes in cut-off values can improve the specificity without much decrement to the sensitivity.

concern for subject confusion, the envelope questions all required a “yes” response whereas half of the neutral and disk questions were answered with “no” responses. This could have changed the cognitive demand resulting in a lack of activation in our regions of interest for the No-crime group. The significantly faster reaction times of the No-crime group versus the Mock-crime group provides some support of the testing being less cognitively demanding for the No-crime group.

Another concern was that the No-crime group did not perform any task and thus was *lying about performing a task* versus *lying about not performing a task* for the Mock-crime group. This was instituted so that a lie condition could be introduced and all operators and investigators doing the analysis would still be blind to the randomization group. If the No-crime subject had picked up the envelope and then reported that he/she did not pick up the envelope, then the researchers performing the scanning would know to which group the participant was randomized. The significantly different group analysis results of *Lie-minus-True* for the Mock-crime and No-crime groups supports, but it does not prove, the idea that the brain activation for lying about doing a task versus lying about not doing the task may be different. Significantly, the Mock-crime

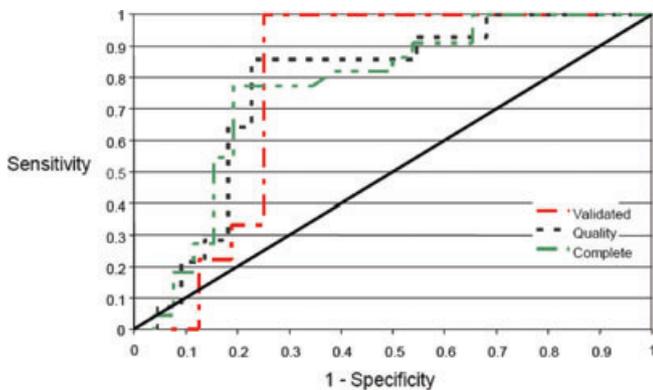


FIG. 3—ROC curves for fMRI detection of deception. Plots of sensitivity and specificity were used to generate curves representing the number of significantly activated voxels at a t -value greater than or equal to 1.645 in clusters 1, 2 and 4 (Kozel et al. 2005) on the task [(disk – neutral) – (envelope – neutral)] for the Complete, Quality and Validated groups. The straight line represents a diagnostic test that is no better than chance and the jagged line represents the trade-off between sensitivity and specificity amongst each group.

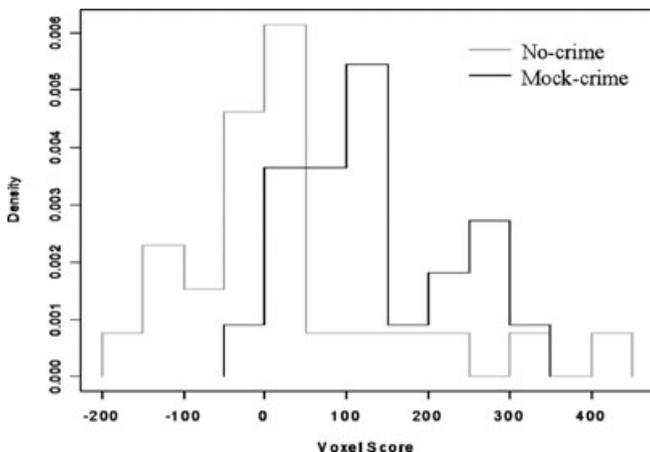


FIG. 4—Distribution of voxel number difference values for the Mock-crime and No-crime participants. The individual participant's voxel difference score for the Complete Group was plotted to assess differences in distribution of values for the Mock-crime and No-crime participants. The x -axis is the voxel difference score and the y -axis is the normalized frequency. This is indicated as "Density" such that the area under the curve for each group is 1. This enables the two groups to be more equitably compared. The Mock-crime (in red) has a distribution clearly to the right of zero (i.e., a correct call would be greater than zero), where as the No-crime (in green) group (i.e., a correct call would be less than zero) has a much more dispersed distribution as would be expected from the sensitivity results and ROC curves.

group whose deception was reporting that they did not do something that they actually did had very accurate individual results using our methodology and a group analysis revealing considerable consistency with our prior study in which subjects also lied about not doing a task. Conversely, the No-crime group whose deception was reporting that they performed a task that they did not perform had much lower accuracy rates and a group analysis that was not very consistent with either the Mock-crime group or the group from our prior study. Future work needs to address these concerns in order to move this technology into practice.

An unexpected finding in this study was the lower rate of accuracy that we found in our Ring-Watch paradigm for this study (71%) versus our prior two groups (93% and 90%) using a similar

paradigm (5). In addition, unlike for the prior two groups that we studied, the group imaging analysis map failed to find a significant result for the *Lie-minus-True* contrast. We suspect that there are a number of reasons for the discrepancy between the results from our prior study and this one. In the prior study, the testing in the scanner for the Ring-Watch paradigm was after a brief motor paradigm scan and participants were given a \$50 incentive if they were able to "fool" a researcher observing the scan. In this study, the testing in the scanner was performed after already being tested for the main focus of study which was the Mock-crime paradigm. Also, there was no incentive given for "fooling" a researcher watching the scan. The potential fatigue of being in the scanner for almost an hour, the reduced salience of the Ring-Watch paradigm, and the lack of a motivating factor may have impacted the results of the methodology. Another concern is that the questioning paradigm was shortened by removing one category of the questions from the previous study. The comparison questions that involved admitting or denying minor offenses were removed because they were not used in the analysis and added 4 min to each scan. The removal of these questions could have reduced the power of the paradigm to detect differences in deceptive versus truthful answers. Successful replications of the original results accomplished at two independent sites supports the idea that differences in accuracy seen were the result of paradigm design differences (data presented as a poster at the American College of Neuropsychopharmacology, December 2007). Both replications used all the questions and gave subjects a monetary incentive to successfully deceive the investigators. Further study will be required to determine the degree of impact that fatigue, salience, motivation, and question format might have, if any, on fMRI detection of deception.

We also compared those participants for whom we made correct classifications and those for whom we made incorrect classifications in their respective groups (Mock-crime and No-crime) to determine if there were demographic or imaging results differences. There were no significant differences in years of education, age, handedness, employment, race, sex, ring or watch taken, time from group assignment to scanning, or number of significantly activated voxels in the ROIs for the ring-minus-neutral or watch-minus-neutral contrasts (data not shown). These factors do not appear to exert an important influence on this methodology. This study was not, however, designed specifically to test these factors.

This study has several factors that must be considered for adequate interpretation of the results. Although this study attempted to approximate a scenario that was closer to a real-world situation than prior fMRI detection studies, it still did not equal the level of jeopardy that exists in real-world testing. The reality of a research setting involves balancing ethical concerns, the need to know accurately the participant's truth and deception, and producing realistic scenarios that have adequate jeopardy. In addition, this study only involved healthy adults who were not taking any medications. Thus, whether fMRI deception testing would work is unknown for participants who are taking a medication, who have a significant psychiatric or medical condition, or who are outside the 18–50 year age range. Future studies will need to be performed involving these populations.

This study has a number of strengths that should be highlighted in the context of how fMRI technology can be moved towards a practical application. The paradigm resembled a real-world scenario for sabotage. The behavior of the participants was closely monitored with sophisticated video equipment and any deviation from the required tasks resulted in participants being excluded. This ensured that we could accurately assign each participant to the Mock-crime or No-crime groups. Data analysis was carried out

TABLE 4—Areas of significant activation for Lie-minus-True group analysis of Mock-crime and No-crime groups.

Cluster	<i>k</i>	Voxel <i>z</i> -value	MNI Coordinates of Voxel with Largest <i>t</i> -value X, Y, Z (mm)	Anatomic Location of Voxel with Largest <i>t</i> -value	Brodmann's Area of Voxel with Largest <i>t</i> -value	Complete Anatomic Area of Cluster
<i>Mock-crime Group Analysis of Lie-minus-True (FDR < 0.05, k > 25)</i>						
1	920	5.29	-45, 30, -9	L orbitofrontal	47	L orbitofrontal
		4.69	-42, 12, 45	L precentral	9	L superior temporal pole
		4.66	-54, 30, -3	L orbitofrontal	45	L insula L precentral L inferior frontal L middle frontal
2	104	4.83	54, 15, -12	R superior temporal pole	38	R superior temporal pole
		3.81	33, 18, -15	R insula	48	R insula
		3.4	36, 27, -6	R orbitofrontal	47	*R orbitofrontal
3	234	4.66	-42, -51, 48	L inferior parietal	40	L inferior parietal
		4.53	-60, -51, 33	L supramarginal	40	L supramarginal
		3.58	-33, -42, 51	L inferior parietal	40	L angular
4	42	4.47	9, -3, 6	R thalamus	NA	R thalamus
		3.94	15, 0, 18	R caudate	NA	R caudate
5	367	4.31	9, 18, 54	R supplementary motor	6	R supplementary motor
		4.03	-3, 18, 48	L supplementary motor	32	L supplementary motor
		3.98	-6, 15, 57	L supplementary motor	6	*R anterior cingulate R middle cingulate L middle cingulate L superior medial frontal R superior medial frontal
6	65	4.23	9, -105, 6	R calcarine	17	R calcarine
		3.38	6, -93, 3	L calcarine	17	L calcarine R superior occipital R cuneus
7	187	4.19	27, -87, -9	R inferior occipital	18	R inferior occipital
		4.08	12, -84, -18	R cerebellum	18	R cerebellum
		3.95	-6, -81, -18	L cerebellum	18	L cerebellum R lingula L lingula R fusiform
8	62	4.01	39, 54, 12	R middle frontal	46	*R middle frontal
		3.81	33, 51, 21	R middle frontal	46	R superior frontal
9	116	3.8	-27, 51, 24	L middle frontal	46	L middle frontal
		3.46	-33, 60, 15	L middle frontal	10	L inferior frontal
		3.42	-45, 51, 6	L middle frontal	46	L superior frontal
10	35	3.7	-18, -102, 12	L middle occipital	17	L middle occipital
						L superior occipital
11	31	3.63	-9, -27, 18	L thalamus	NA	L thalamus L hippocampus
12	50	3.51	-24, -63, 60	L superior parietal	7	L superior parietal
		3.32	-15, -69, 60	L precuneus	7	L precuneus
13	34	3.28	-60, -36, -12	L middle temporal	20	*L middle temporal
		3.2	-51, -42, -6	L middle temporal	21	L inferior temporal
<i>No-crime Group Analysis of Lie-minus-True (FDR < 0.05, k > 25)</i>						
1	576	5.02	-3, -24, 48	L middle cingulate	NA	L middle cingulate
		4.39	-12, -18, 42	L middle cingulate	NA	R precuneus
		4.19	9, -54, 33	R precuneus	23	R calcarine L calcarine L precuneus L cuneus R posterior cingulate L posterior cingulate L supplementary motor R supplementary motor L paracentral lobule
2	659	5.01	12, 54, 0	R orbitofrontal	10	*R orbitofrontal
		4.97	-12, 42, -15	L rectus	11	L rectus
		4.87	-3, 48, -18	L rectus	11	L orbitofrontal R rectus L anterior cingulate *R anterior cingulate L medial superior frontal R medial superior frontal
3	29	4.64	-45, -78, 36	L middle occipital	39	R superior frontal L middle occipital L angular
4	224	4.42	45, -27, 63	R postcentral	3	R postcentral
		3.94	42, -18, 45	R postcentral	4	R precentral

TABLE 4—(Continued)

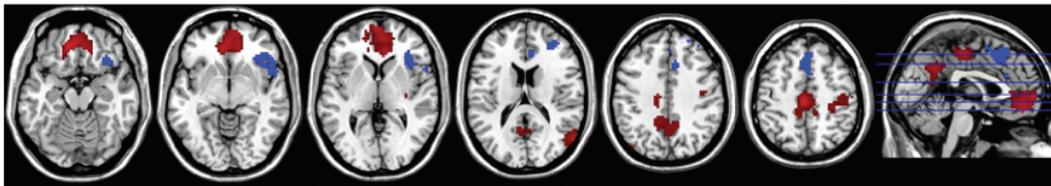
Cluster	<i>k</i>	Voxel <i>z</i> -value	MNI Coordinates of Voxel with Largest <i>t</i> -value X, Y, Z (mm)	Anatomic Location of Voxel with Largest <i>t</i> -value	Brodmann's Area of Voxel with Largest <i>t</i> -value	Complete Anatomic Area of Cluster
5	117	3.86	30, -30, 54	R postcentral	3	R middle temporal R middle occipital R superior temporal
		4.22	57, -66, 18	R middle temporal	39	
		3.84	42, -51, 21	R middle temporal	39	
6	27	4.15	30, -12, 9	R putamen	48	R angular R putamen R insula
7	27	3.76	21, -18, -27	R parahippocampus	30	R parahippocampus
		3.46	27, -12, -27	R parahippocampus	36	

*Regions that have been significantly activated in our prior 3 studies for Lie-minus-True.

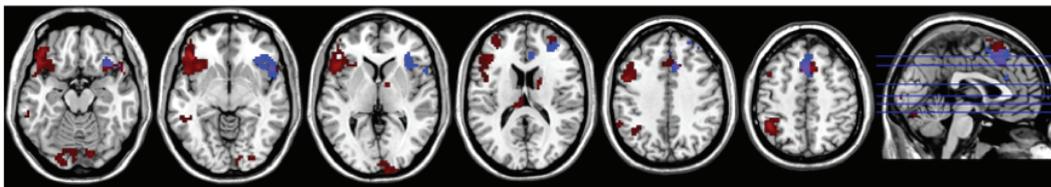
R, right; L, left; *k*, number of voxels in cluster; NA, location does not have a Brodmann's Area value such as being located in deep brain structures or white matter.

Brodmann's Areas and anatomic locations were determined using MRICro and finding the closest reasonable value.

No-crime Group Map of Lie-minus-True, *n* = 22 (FDR<0.05, *k*>25)



Mock-crime Group Map of Lie-minus-True, *n* = 14 (FDR<0.05, *k*>25)



Model Building Group – Kozel et al. 2005 of Lie-minus-True, *n* = 30 (FDR<0.05, *k*>25)

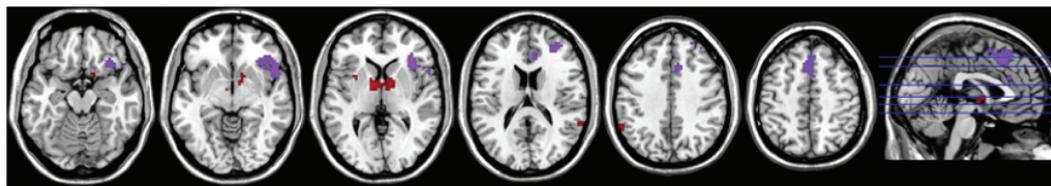


FIG. 5—Group results for Mock-crime and No-crime groups (participants meeting quality requirements). Group analysis of Lie-minus-True for three independent groups of subjects: the Mock-crime group; the No-crime group; and the Model Building group from our prior fMRI detection of deception study (5). This Model Building Group was used to determine the regions of interest and methodology for this analysis. The axial images of the structural template in MRICron (<http://www.sph.sc.edu/comd/rorden/mricron/>) have areas of significant activation (FDR < 0.05, *k* > 25) for the respective groups overlaid in red. The voxels corresponding to the a priori determined regions of interest used in the analysis are indicated with blue. The voxels in which both the activation and regions of interest overlap are indicated with purple. The sagittal images to the right of the figure specify the levels of the axial slices. The *z* levels for all groups are -16, -6, 3, 17, 39, and 50 from left to right of the image. The right of the axial brain image is the right side of the brain. Similarities can be seen between the activation pattern in the Mock-crime group and the Model Building group. Conversely, the No-crime group and the Model Building group have quite differing activation patterns.

blind to the participant assignment in a largely automated manner. The investigators, who performed the analysis blind to group assignment and base rate of how many participants would be asked to commit the mock-crime, used *a priori* defined analysis methods run with Matlab scripts that ensured repeatability of the methods. The only operator judgment required was the assignment of the 0,0,0 coordinate to the location of anterior commissure on the BOLD fMRI image. Despite two operators performing this

independently and with differing results, the final analysis outcomes were remarkably similar. In 94 of 96 cases both researchers drew the same conclusions. In the two discordant cases, the researchers could not draw a conclusion because of disagreement of results. The results between the two analyses, however, were close. One of these two participants was found to have a result of 0 activated voxels (indeterminate call) for one investigator and 1 activated voxel (committed mock-crime call) for the other investigator, while

the other participant had the investigators getting the inverse results (1 activated voxel—committed mock-crime call, 0 activated voxels—indeterminate call). The close agreement of the two independent analyses highlights the robustness of the data analysis methods. Finally, though our *a priori* hypothesis resulted in low specificity and high sensitivity, the ROC analysis shows that the test may be tailored to the situation where one is investigating.

Our methodology of using functional MRI to detect deception was found to be sensitive but suffers from low specificity on this task for whether a subject committed a mock crime. This indicates that the test would be helpful to “rule out” a potential suspect (i.e., a person who is found to be not lying about being innocent—did not commit the crime) but not very helpful in “ruling in” a suspect (i.e., a person who is found to be lying about being innocent—did commit the crime). Comparing these results with other testing modalities is problematic due to the variability in testing and groups of people tested. The most comprehensive report on the literature surrounding the use of the polygraph to detect deception simply concluded that although there was a considerable variability in results, the polygraph was probably better than chance (25). More work with direct comparisons of paradigms and participant samples are needed to understand how the various technologies compare in detecting deception. Although the diagnostic ability of our method was greater than chance, future work is focused on improving specificity and using more realistic testing in order to enhance the utility of this technology in real-world applications.

Funding

Funding provided by Department of Defense Polygraph Institute (W74V8H-04-1-0010) and Cephos Corp. Dr. Kozel is supported by a K23 from the NIMH (5K23MH070897-01). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of Defense, the National Institute of Mental Health, the National Institutes of Health, or the U.S. Government.

Acknowledgments

We wish to thank Dr. Margaret Melikian, Director of Forensic Psychiatry at MUSC, for the use of the specially equipped room in the Forensic Psychiatry Building; Mr. Dave Ramsey and Mr. John Dornisch of the South Carolina Research Authority for computing support; Drs. Paul Morgan and Chris Rorden for help with imaging parameters; and Ms Minnie Dobbins and Ms Kimberly Mapes for administrative support.

References

- Lee TM, Liu HL, Tan LH, Chan CC, Mahankali S, Feng CM, et al. Lie detection by functional magnetic resonance imaging. *Hum Brain Mapp* 2002;15(3):157–64.
- Langleben DD, Loughhead JW, Bilker WB, Ruparel K, Childress AR, Busch SI, et al. Telling truth from lie in individual subjects with fast event-related fMRI. *Hum Brain Mapp* 2005;26(4):262–72.
- Ganis G, Kosslyn SM, Stose S, Thompson WL, Yurgelun-Todd DA. Neural correlates of different types of deception: an fMRI investigation. *Cereb Cortex* 2003;13(8):830–6.
- Kozel F, Revell L, Lorberbaum J, Shastri A, Nahas Z, Horner M, et al. Brain regions involved in deception: an fMRI study in healthy volunteers. *J Neuropsychiatry Clin Neurosci* 2001;13:147 (Abstract).
- Kozel FA, Johnson KA, Mu Q, Grenesko EL, Laken SJ, George MS. Detecting deception using functional magnetic resonance imaging. *Biol Psychiat* 2005;58(8):605–13.

- Spence SA, Farrow TF, Herford AE, Wilkinson ID, Zheng Y, Woodruff PW. Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* 2001;12(13):2849–53.
- Langleben DD, Schroeder L, Maldjian JA, Gur RC, McDonald S, Ragland JD, et al. Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* 2002;15(3):727–32.
- Spence SA, Hunter MD, Farrow TF, Green RD, Leung DH, Hughes CJ, et al. A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philos Trans R Soc Lond B Biol Sci* 2004;359(1451):1755–62.
- Kozel FA, Revell LJ, Lorberbaum JP, Shastri A, Elhai JD, Horner MD, et al. A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *J Neuropsych Clin N* 2004;16(3):295–305.
- Kozel FA, Padgett TM, George MS. A replication study of the neural correlates of deception. *Behav Neurosci* 2004;118(4):852–6.
- Mohamed FB, Faro SH, Gordon NJ, Platek SM, Ahmad H, Williams JM. Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation—initial experience. *Radiology* 2006;238(2):679–88.
- Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughhead JW, et al. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage* 2005;28(3):663–8.
- Nunez JM, Casey BJ, Egner T, Hare T, Hirsch J. Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage* 2005;25(1):267–77.
- Phan KL, Magalhaes A, Ziemlewicz TJ, Fitzgerald DA, Green C, Smith W. Neural correlates of telling lies: a functional magnetic resonance imaging study at 4 tesla. *Acad Radiol* 2005;12:164–72.
- First M, Spitzer R, Williams J, Gibbon M. Structured clinical interview for DSM-IV (SCID). Washington, DC: American Psychiatric Press, 1995.
- Annette M. A classification of hand preference by association analysis. *Br J Psychol* 1970;61:303–21.
- Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the state-trait anxiety inventory. Redwood City, CA: Mind Garden, Inc, 1983.
- Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. *Arch Gen Psychiatry* 1993;50(12):975–90.
- Kircher JC, Raskin DC. Human versus computerized evaluations of polygraph data in a laboratory setting. *J Appl Psychol* 1988;73(2):291–302.
- Hintze J. NCSS, PASS, and GESS. Kaysville, Utah: NCSS, <http://www.ncss.com>. Accessed October 31, 2008.
- Bullmore E, Brammer M, Williams SC, Rabe-Hesketh S, Janot N, David A, et al. Statistical methods of estimation and inference for functional MR image analysis. *Magn Reson Med* 1996;35(2):261–77.
- Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* 1999;7(4):254–66.
- Friston K. Experimental design and statistical parametric mapping. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, et al., editors. *Human brain function*, 2nd ed. Amsterdam: Elsevier Academic Press, 2004;599–634.
- Friston KJ, Frackowiak RSJ. Images of the future. In: Frackowiak RSJ, Friston KJ, Dolan RJ, Frith CD, Mazziotta JC, editors. *Human brain function*. San Diego, CA: Academic Press, 1997;487–517.
- National Research Council. The polygraph and lie detection. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 2003.

Additional information and reprint requests:

F. Andrew Kozel, M.D., M.S.C.R.

Assistant Professor

UT Southwestern Medical Center

Department of Psychiatry

5323 Harry Hines Blvd.

Dallas, TX 75390-9119

E-mail: andrew.kozel@utsouthwestern.edu