Determination of Truth from Deception Using Functional MRI and Cognitive Engrams
D Marks, M Adineh, S Gupta

Abstract

INTRODUCTION

Investigators have attempted for centuries to determine the truthfulness and accuracy of statements (Ford 2006). Methodologies include standard interrogation, with and without use of medication or coercion, polygraph, and brain fingerprinting using brain evoked response potentials (ERP). None of these technologies yields optimal results, for various reasons, and are not entirely suitable for measuring deception.

An association between the ERP and lying on a standard test such as the Guilty Knowledge Test (GKT) suggested that deception may be associated with changes in other measures of brain activity such as regional blood flow (RBF) (Langleben 2002; Mochizuki 2001). Changes in RBF, as Blood Oxygen Level Dependent (BOLD) can be monitored using functional magnetic resonance imaging (fMRI), and have been shown to be measurements of regional brain activity (Ogawa 1990). These areas of brain activation on fMRI have been correlated with active memory encoding (Wegner 1999).

A number of researchers (Slotnick et al, Langlaben et al, Lee et al, Kozel et al) have sought to apply fMRI to discern between truth and deception. In general, test subjects were presented with artificial situations during which they were given the opportunity to tell the truth or give a false response to questions, while undergoing fMRI. The design of presentation of questions were different between research groups. These separate research groups were able to detect truthful or deceptive responses, based upon analysis of activation maps showing where in the brain there was increased neural activity. The areas of activation corresponding to truth or deception differed among groups of researchers, making a universal framework for analysis of interrogation with fMRI difficult to develop.

The purpose of this study was to validate prior work of other investigators, and to develop a paradigm to address inter-investigator variability that needs to be resolved before fMRI can be applied to interrogation in practice. Our methods employ activation area analysis, both for truth and deception, and also use activation map patterns for object recognition. These combination analytical methods should yield a more robust schema for use of fMRI as an interrogation technique. One of the central means of analysis we employ is the concept of the cognitive engram, which we define as the multi-dimensional representation of activation areas in the brain, which we contend represent the actual specific thought process. We have grouped truthful and deceptive responses, and also the recognition of specific faces and objects, into cognitive Engrams. Certainly when an individual thinks of a truthful or a deceptive response, or recognizes a face or an object, that concept transforms into the more generalized concept of truthfulness or deception, or recognition of a specific person. This overall recognition is not simply fragmented into hundreds of isolated activation points, but should be treated as a whole, as a cognitive engram, consisting of a number of widely distributed activation areas representing a single concept.

METHODOLOGY

We operated under an IRB-approved protocol, and all volunteers signed an informed consent form. Healthy adult volunteers without exclusions to MRI scanning were explained the study purpose and design, risks and benefits, given a chance to ask questions and then agreed to participate. Continual neuroimaging (fMRI) was performed during viewing of the test stimuli in order to capture structural and functional data. The data was analyzed for the presence of neuroimaging activation that has been shown to correspond to cognition and visual recognition.
TEST STIMULI
Two different visual stimuli in a form of color photographs were generated and presented using Microsoft Powerpoint. The visual stimuli were projected onto a flat panel rear projection display placed at the foot of the test subjects, who viewed the screen by a mirror system attached to the head coil.

Figure 1
Figure 1: Visual stimulus 1

Figure 2
Figure 2: Visual stimulus 2

DATA ACQUISITION
A 3-T General Electric Signa 11X Excite scanner with whole head coil was used.

A short localizer MRI scan was performed to verify that the field of view was within the skull, and that there were no “ghost” images. A high-resolution full volume structural MRI scan was then obtained for each subject, using fast SPGR imaging (146, 1.0-mm thick axial slices, no spaces, TR = 8, TE = 3.2, FOV = 24 cm, 256 x 256 matrix). These T1-weighted images provided detailed anatomical information for registration and 3-D normalization to the Talairach and Tournoux atlas (1988).

Changes in the blood oxygen level dependent MRI signal were measured using a gradient-echo echoplanar sequence. Functional MRI (fMRI) scans lasted 110 seconds each. EPI parameters were: TE 35, TR 2000, multiphase screen, 55 phases per location, interleaved, flip angle 90, delay after acquisition-minimum. Using a visual stimulus package, color photographs were presented in a mini-block design. In a typical session, after a 4 second lead-in time, a blank screen was displayed for 4 seconds, then the picture of interest for 4 seconds, and this was repeated for the scan time.
There were two separate runs for each test visual stimulus, and during the first run each test subject was instructed to think that they recognized that visual stimulus (the truthful response). During the second test run, the test subject was shown the second visual stimulus, and instructed to think that they recognized him (a second truthful response). For the third run, each test subject was shown the first visual stimulus, and instructed to think that he did not recognize that person (false or deceptive response). For the last run, each subject was shown the second visual stimulus, and instructed to think that they did not recognize him (a second false response).

**CREATION OF ACTIVATION POINTS**

The fMRI scan volumes were motion-corrected and spatially smoothed in-plane. MRI data files were normalized and analyzed using MedX (version 3.4.3, Sensor Systems, Sterling, VA) to compute statistical contrasts and create a map representing significantly activated areas of the brain that responded differentially to a visual test stimuli.

For the voxels that show an overall increase in activity for meaningful stimuli, a positive regression analysis for the contrast between a test photo and control (blank page) stimuli was conducted, creating an activation map containing specific voxels with an uncorrected probability, P ≤ 0.05; meaning every voxel showing activation with the probability greater than 95%. Only those voxels were selected for further analysis. That statistical map was then superimposed on coplanar high-resolution structural images. The partial volume structural images were registered with the full volume high-resolution images using Automated Image Registration (Woods 1993).

Those full volume high-resolution images were then transformed (registered and normalized) to the Talairach and Tournoux atlas (1988) using MedX tools. Each activated voxel on these images was selected to obtain Talairach coordinates of brain regions that respond maximally to the test stimuli and to further generate a Cognitive Engram. Comparison of observed patterns of activation were correlated with the nature of the response: Truthfulness or Deception.

Three dimensional graphical representations of the identified activation maps were constructed by plotting the xyz coordinates, using the program DPlot (HydeSoft Computing, Vicksburg, MS).

**RESULTS**

Five normal individuals were interrogated on whether they recognized the two test visual stimuli, and instructed to tell the truth. All 5 persons did in fact recognize both visual stimuli before undergoing MRI, and their thoughts/intentions at that time were truthful. Activation points were identified for each individual tested, and those activation points present in at least four of five test subjects viewing either test photo were termed consensus points, truth. The coordinates for these fMRI activation areas are listed in Table 1, as Marks 2006, Truth>False (T>F), and they are graphically displayed in Figure 3 (for visual stimulus #1) and for Figure 5 (for visual stimulus #2).

**Figure 3**

Table 1: Activation areas from normal test subjects across investigators during mock interrogation episodes.
These same five normal individuals were then interrogated on recognition of the two test visual stimuli, and were instructed this time to think “non-recognition” while viewing the photos. As all 5 persons did in fact recognize both test visual stimuli, their thoughts/intentions at that time were false/deceptive. Activation points were identified for each individual tested, and those activation points present in at least four of five test subjects viewing either test photo were termed consensus points. The coordinates for these fMRI activation areas representing a False/deceptive response are listed in Table 1, as Marks 2006, F>T, and they are graphically displayed in Figure 4 for visual stimulus #1 and in Figure 6 for visual stimulus #2.

Legend for Figures 3 and 4: fMRI activation maps of individuals who viewed visual test stimulus #1 while undergoing fMRI. The first time they viewed the photo, they were asked if they recognized the person, and to be truthful. The second time they were instructed to think that they did not recognize that person (false response). Shown are 3-D scatter point graphs, Figure 3 (truth) and Figure 4 (false/deceptive).
Legend for Figures 5 and 6: fMRI activation maps of 4 individuals who viewed a photo of visual test stimulus #2 while undergoing fMRI. The first time they viewed the photo, they were asked if they recognized the person and to be truthful. The second time they were instructed to think that they did not recognize him (false response). Shown are 3-D scatter point graphs, Figure 5 (truth) and Figure 6 (false/deceptive).

The Truthful response data when viewing photo of both test both visual stimuli were combined, and were presented in Figure 7, as 3-D scatter point graphs. Similarly, the False / deceptive response data for both test visual stimuli were combined, and were presented in Figure 8, as 3-D scatter point graphs.

Legend for Figures 7, 8: The next figures represent the activation points that were in common for visual test stimuli #1 and #2 for a truthful response (Fig. 7), or for a false/deceptive response (Fig. 8). Shown are 3-D scatter point graphs.

Data sets were identified within the T>F and F>T that were not overlapping, and these “consensus” data sets are displayed in Figures 9 and 10. They represent the “minimal”
data sets that could be used to distinguish a truthful from a false/deceptive response.

Figure 12
Figure 9

Legend for Figures 9, 10: Consensus activation points that were common for visual test stimuli #1 and #2 for a truthful response (Fig. 9), that were unique and not overlapping with the consensus points for a false/deceptive response (Fig. 10). Shown are 3-D scatter point graphs, based upon the data points listed in Table 1 (listed for Marks 2006).

We also present data on simple recognition of visual test stimuli #1 and #2, as examples of visual activation for two different human male faces. Test subjects were shown test images 1 and 2 while undergoing fMRI, and no questions or instructions were given to the viewer. The patterns represent the cognitive Engrams for those two faces.

Figure 14
Figure 11

Legend for Figures 11 and 12: Pattern of recognition for visual test stimuli #1 and #2, without asking the viewer if they recognized that person or not. This formed the activation patterns for visual facial recognition of facial test stimuli #1 and #2, their corresponding cognitive engrams. Shown are 3-D scatter point graphs.

DISCUSSION
We have shown that specific activation patterns occur in the
Determination of Truth from Deception Using Functional MRI and Cognitive Engrams

brain of individuals looking at specific pictures, and also whether they are contemplating giving a truthful or a deceptive response. We hold that the consensus activation points across test subjects viewing a photo of visual face test stimulus #1 (Fig. 11) and #2 (Fig. 12) represent symbolically cognitive engrams, which may represent or correlate to the actual concept of visual face test stimulus as a thought. That pattern of visual recognition is a separate construct from the activation map for questioning on recognition of a visual face stimulus, which involves facial recognition, but also includes the direction to tell the truth with regards to recognition of that individual. Thus Figures 3 and 5 may represent the sum of a visual recognition for those visual face test stimuli along with the truthful response to whether they are recognized.

The cognitive engram patterns for visual face test stimulus #1 (Fig. 11) and for visual face test stimulus #2 (Fig. 12) are distinct and non-overlapping, facilitating the determination of which photo a person was viewing (and thinking about) by examining the activation pattern in their brain via fMRI. It could be possible to create a library of visual images and their corresponding cognitive Engrams, to allow the interpretation of activation patterns without knowing in advance the stimulus.

The patterns for visual face test stimulus #1 truth (Fig. 3) vs. #1 false/deceptive (Fig. 4) were unique and non-overlapping, confirming in our model that it is possible to distinguish a truthful from a deceitful response via functional MR imaging. The same conclusions can be reached looking at activation patterns for Truth and False when viewing the visual face test stimulus #2 photo (Figures 5 and 6 respectively). That the activation patterns for visual face test stimulus #1 – truth (Fig. 3) and visual face test stimulus #2 – truth (Fig. 5) are not overlapping probably reflects the different components of a truthful response: visual recognition, emotional overlay, activation of or suppression of areas in the brain involved with a truthful response or a false/deceptive response, the later which may be at least in part the suppression of a truthful response, and the design of the interrogation. These and other factors (language, culture, medication, disease state, age, etc.) must all be factored into any system which purports to apply fMRI to interrogation. This is especially important when the outcome of an interrogation can have serious consequences to the person undergoing interrogation.

The combined data (Fig. 7) for both visual face test stimulus #1 and #2 when looking for recognition, with a truthful response, were not overlapping or similar to that for a deceptive response (Fig. 8). Similar results were seen with consensus data (Fig. 9, 10) in addition to combined data. This supports our approach to determining whether an activation map was indicative of a truthful or a deceptive response, by looking at consensus activation maps.

The three dimensional plots of activation maps may represent symbolically the actual concept of guilt / deception in the processing and memory areas of the brain. In this way, the maps generated here describe Cognitive Engrams (CE), or data representations of concepts, in this case truth or deception. In this three dimensional connectiveness, CE may be in some ways analogous to the Object Form Topography (OFT) described by other researchers (Haxby 2001). However, Cognitive Engrams differ from OFT in that CE appear to be the actual symbolic representation (Fig. 4-12) of actual thought concepts (persons, places, things, intents) in this case, the concept of truth v deception, and recognition of specific faces of well-known men. Each CE was given a unique identifier detailing the concept represented and the source of the data.

Interrogation for the determination of truth must sort out deceptive or inaccurate responses. Not all individuals attempting to respond truthfully may remember the details they are being questioned about, because of incomplete remembrance of all details of an event, and the normal deterioration or fading of memory with time. There may also be confusion on what was being remembered, or perceived to being remembered, so-called reality monitoring errors (Johnson 1993; Dodson 1993). Mistaken, or false, remembering also plays a role in the inaccuracies of interrogation. Functional neuroimaging has shown, not unexpectedly, that words well-remembered as opposed to words later forgotten show greater activity during the act of remembering (encoding) in those areas of the brain associated with active remembering (for ex., left inferior prefrontal cortex (Reber 2002).

Gonsalves et al (2004) used neuroimaging to study false remembering. In their experiment, normal subjects were shown a series of pictures of a noun typed onto a display, and told to form a visual image of the subject of each word as it appeared. This model was known to lead to some reality monitoring error. Half of the graphical displays of individual words were followed by a picture representing the word, for example, the word CAT, followed by a photo of a cat. Half of the words were not followed by a photo representing the word. Functional MRI was conducted during the testing, and
a structural MRI scan followed to correlate functional imaging with anatomic neuroanatomy. After the viewing of cards with words typed on them, and outside of the MRI scanner, individuals were presented randomly with 525 words, but this time spoken, and asked if a picture had been previously presented for each word. One hundred seventy five words had been previously presented followed by a photo while undergoing neuroimaging, 175 had been presented but not followed by a photo and 175 were new nouns, not presented previously. Correct identification of having seen a photo with a printed word was 74%, falsely claiming to have seen a photo after a given displayed word was 27%, and 6% of subjects claimed to have seen a photo for a word that had as not even been presented (false alarm). Gonsalves et al found that activations in two areas of the brain (precuneus and inferior parietal regions of cerebral cortex), known to be active in visual images, distinguished true from false remembrance, indicating that the brain activity engaging visual imagery can lead to falsely remembering something only imagined. For comparison, we saw activation in the inferior parietal and cuneus but not precuneus regions.

Langleben and colleagues (2002) studied fMRI contrasts between deceptive and truthful responses. Eighteen participants were measured with a 4 Tesla MRI scanner while performing a “high motivation” modification of the GKT, using a modified deck of playing cards, and analyzed using statistical parametric mapping. The GKT facilitates psychophysiological detection of prior knowledge of crime details that would be known only to a suspect involved in the crime (Lykken 1991). They modified the GKT in ways to minimize anxiety response, while maintaining the motivation to deceive with modest positive reinforcement. In their model, increased activity in the anterior cingulate cortex (ACC), the superior frontal gyrus (SFG), and the left premotor, motor, and anterior parietal cortex was specifically associated with deceptive responses, indicating that cognitive differences between deception and truth have neural correlates detectable by fMRI. These findings are not consistent with our own findings. Although Langlaben et al (2002) found brain regions that were more active during Lie than Truth (Figure 13), there were no regions more active during Truth than Lie, suggesting that the truth is the baseline cognitive state, but certainly limits the use of these findings to detecting deception as opposed to verifying the truthfulness of a response. In our own case, and those of other researchers, were able to distinguish a truthful from a deceptive response. As expected from Langleben et al
As was the case for the research subjects used by other groups (Langlaben, Slotnick), the volunteers studied by Kozel et al (2004, 2005) were devoid of a wide range of conditions potentially confounding interpretation of fMRI in real life interrogation circumstances (extreme stress, ideologic commitments, underlying illnesses, use of prescription and illegal drugs with CNS effects, etc.). The fMRI results showed that for lying, when compared with telling the truth, there was more activation in at least five different brain regions (identified as Kozel 2004 in their Table 2, and reproduced in our Table 1). While activation during deception was detected with fMRI in the orbitofrontal cortex (OFCx) and anterior cingulate (AC), individual results were not consistent and lacked good predictive power for individuals. We did not find activation in these two areas (Table 1). As was the case for the work of Langlaben et al, Kozel et al (2004) did not find their technique capable of detecting a truthful response, only for picking a signal indicating that deception was being employed, and they concluded that at that time, fMRI could not be used for that purpose.

In a follow-up study, Kozel et al (2005) again used fMRI, this time in a mock crime paradigm, and the data analyzed with as a model-building group. This yielded a set of activation areas (identified as Kozel 2005 in our Table 1) that could correctly differentiate a truthful from a deceptive response. This modeled data was then used to correctly identify truth/deception in an independent group of individuals, using the same mock crime paradigm. Interestingly, there are no uniform, consistent major anatomical overlap sets (by Talairach coordinates or Brodmann areas) for brain activation areas associated with truth and deception questioning when comparing the results from Kozel 2004 to Kozel 2005 (Table 1). Nor did there appear to be a common area of consensus activation noted when the data are plotted out in three dimensions (Figure 7, 9, 13-15). Certainly part of the differences are due to the experimental models, the questions asked, or variations in subjects and in imaging techniques. However, without overlap or common consensus areas between researchers or even consistently within one research group it would seem that no one set of data identified by any one author can reliably, definitively, or conclusively alone be used to discern truth / deception. This could indicate either that the exact model used for interrogation is crucial, or that there is a problem with internal consistency. Any real life application will require the model to be robust enough to work across individuals, questioning techniques and imaging paradigms.

Legend for Figure 14: Activation points shown by Slotnick (2004) to distinguish a Truthful from a False response.

Kozel (2004)
LIMITATIONS OF FUNCTIONAL MRI FOR DISTINGUISHING TRUTH FROM DECEPTION

The experiments (Kozel; Langleben and others) used to generate these data and identify these activation sites were prepared from research protocols that were, by their very nature, designed to maximize positive results. Experienced liars, criminals, and individuals taking mind-effecting pharmaceuticals (serotonin reuptake inhibitors, amphetamines, cocaine, etc.) were not included. The experimental models were not tested in real life circumstances, and their application to such conditions is not yet established.

All experiments for which truth / false cognitive engrams were constructed (Kozel, Langleben, Slotnick) used small numbers of normal (drug free, non-criminal) volunteers, involved highly structured and artificial systems, none of the systems were applied to actual criminal investigations, and there unfortunately appeared to be little correlation between the experimental results between research groups. Large numbers of replicate scans under highly controlled circumstances are always needed in fMRI research to accommodate for intersubject and interscan variability (McGonigle 2000).

Analyses of experimental data by Kozel (2004, 2005), Langleben (2004) and others typically use a minimum cutoff size for number of voxels needed to yield a useful cluster. Although removing spurious data, in an environment of much inter-subject variability, this practice also eliminates potential low level data that could indicate finer specificity and details that code memory and intent for a given person. We tended to keep for analysis as many of the activation points as possible, even if of low voxel size and not in clusters, if the points were consistently activated across individuals.

Lee et al (2002) used fMRI on six healthy male Mandarin-speaking volunteers who underwent functional magnetic resonance imaging with a block-design paradigm while they performed forced-choice memory tasks involving both simulated malingering and under normal control conditions. Malingering that demonstrated the existence and involvement of a prefrontal parietal-sub-cortical circuit with feigned memory impairment produced distinct patterns of neural activation. Wyland et al evaluated neural correlates of thought suppression, and found that suppression of a particular thought, when compared to the free-thought control condition, revealed greater activation in the anterior cingulate. When the task of suppressing all conscious thoughts was compared to free-thought, a more distributed network of brain regions, including the anterior cingulate and the insula, was activated. Certainly the work of Lee et al, and of Wyland et al, gives some direction to the potential superiority of fMRI analysis of truth / deception testing when compared to more conventional technologies, such as the polygraph.

It would appear from our analysis that a greater likelihood for success to distinguish a truthful from a deceptive response during interrogation in a real life situation will take a multifactorial approach to studying brain activation with fMRI. Specifically, fMRI will need to be combined with determination of the presence of specific concepts (cognitive engrams of persons, places, intentions) materially related to the subject of investigation. A concordance between the cognitive engrams for truth / deception, as demonstrated above in Figure 7-10, with the presence of specific cognitive engrams for the matter of inquiry (person, place, intent) may allow much improved determination of whether a response is truthful or deceptive. Further, at the beginning of each fMRI run, a set of known truth and a known false/deceptive responses should be run to “calibrate” for that individual which sets of activation points are best.
CONCLUSIONS

Future fMRI work will need to address normal subjects, persons who are skilled at deception (including criminals, terrorists, and persons who are trained at deception), persons on CNS-active medications (both prescription medications and illicit drugs), the effects of language and culture, and a number of other factors. We have begun to modify our experimental designs to take these factors into account, and we expect that fMRI will successfully be applied to interrogation with a high degree of confidence.

ACKNOWLEDGEMENTS

We are grateful to St. Vincent's Hospital in Birmingham, Alabama for the generous use of their 3T MRI imaging time, and Cooper Green Mercy Hospital for use of their clinical facilities.

References


r-3. Dodson CS, Johnson MK. Rate of false source attributions depends on how questions are asked. Am J Psychol. 1993 Winter;106(4):541-57.


r-13. Langleben DD et al. Telling truth from lie in individual subjects with fast event-related fMRI. Human Brain Mapping 13 Sep 2005


Author Information

Donald H. Marks, M.D., Ph.D.
Wallace-Kettering Neuroscience Institute

Mehdi Adineh, Ph.D.
Wallace-Kettering Neuroscience Institute

Sudeepa Gupta, B.S.
Wallace-Kettering Neuroscience Institute