Detection of deception with fMRI: Are we there yet?

Daniel D. Langleben*
Department of Psychiatry, University of Pennsylvania and the Veterans Administration Medical Center, Philadelphia, USA

A decade of spectacular progress in functional magnetic resonance imaging (fMRI) technology and systems neuroscience research has so far yielded few changes in our daily lives. The dearth of clinical applications of this prolific and academically promising research tool began raising the eyebrows of the public and the research funding agencies. This may be one of the reasons for the enthusiasm and interest paid to the growing body of literature suggesting that blood oxygenation level-dependent (BOLD) fMRI of the brain could be sensitive to the differences between lie and truth. The word 'differences' is critical here since it refers to the often-ignored core concept of BOLD fMRI: it is only sensitive to differences between two brain states. Thus, available studies report using fMRI to discriminate between lie and truth or some other comparative state rather than to positively identify deception. This nuance is an example of the extent to which applied neuroscience research does not lend itself to the type of over-simplification that has plagued the interpretation of fMRI-based lie detection by the popular press and the increasingly vocal academic critics. As an early contributor to the modest stream of data on fMRI-based lie detection, I was asked by Dr Aldert Vrij to write a piece in favour of fMRI-based lie detection, to be contrasted with a piece by Dr Sean Spence presenting an opposite point of view (Spence, 2008). This seemingly straightforward task presented two hurdles: having to respond to the popular as well as scientific view of what lie detection with fMRI is and present a wholly positive view of evolving experimental data.

Deceit is ubiquitous in humans (Vrij, 2001). Recognition of deception in others carries significant survival benefits (Rowell, Ellner, & Reeve, 2006), but is not well developed in normal individuals (Etoff, Ekman, Magee, & Frank, 2000). For that reason, social interactions may be routinely discounted for possible deception (Lachmann & Bergstrom, 2004). Reducing or eliminating the likelihood of deception could alleviate the burden of mistrust and increase the efficiency of human interactions. Hence, every generation has attempted to develop objective and reproducible methods to discover the 'truth', using technology of the age as well as a cast of practitioners of these

*Correspondence should be addressed to Dr Daniel D. Langleben, MD, Treatment Research Center, 3900 Chestnut Street, Philadelphia, PA 19104, USA (e-mail: langlebe@upenn.edu).

DOI:10.1348/135532507X251641
methods (Eck, 1970). In the clinical setting, ‘objective truth’ is a record of past events, reconstructed from physical evidence and human memory. Though the presence of deception points to what the objective truth may be, detecting deception is only an indirect path to the truth through elimination of false leads.

The last two centuries of progress in the natural sciences have provided an alternative to torture and other ‘empirical’ aids to lie detection (Trovillo, 1939), and ushered in the development of psychophysiological technique, commonly referred to as the polygraph (Matte, 1996; Kleiner, 2002; Stern, 2004). The physiological component of the polygraph tracks multiple measures of activity of the peripheral nervous system (skin conductance, perspiration, blood pressure, and breathing patterns). The ‘psychological’ component is the form of questioning itself, which constricts the exchange of information between examiner and examinee to a forced-choice format more amenable to formal analysis. Example of such format is the ‘Comparison Question Test’ (CQT; Stern, 2004). Two of the many real and imaginary failings of the polygraph are of particular importance: the interpersonal expectancy (i.e. bias) effect of the examiner/examinee interaction (Rosenthal, 1994) and the peripheral rather than central nervous system source of the physiological measures of the polygraph which places a ceiling on the accuracy potential of this technique (Stern, 2004). Nevertheless, the polygraph remained the mainstay of forensic lie detection for almost a century.

The introduction of electroencephalography (EEG) as the new means of lie detection (Farwell & Donchin, 1991; Rosenfeld et al., 1988) did not dislodge peripheral physiological measures from their pre-eminent position, but it provided a glimpse of the potential advantages of central neurophysiologic measures in assessing processes that take place in the brain. EEG or, more precisely, event-related brain potentials (ERPs) were combined with the Guilty Knowledge Test (GKT), another common deception-generating paradigm that is radically different from the CQT. The GKT is best described as a ‘prior knowledge test’, where items of higher salience to the participant, regardless of cause, generate larger ERP response (Farwell & Donchin, 1991; Farwell & Smith, 2001). Thus, unless a GKT is controlled for item salience due to causes other than deception, it may be unable to discriminate between a lie and an item, salient for reasons other than deception. Indeed, intentionally increasing the salience of non-lie items has been a successful countermeasure to ERP/GKT lie detection (Rosenfeld, 2004). Though GKT is also used for psychophysiological lie detection, the purists reserve the term ‘polygraph’ to psychophysiological record of a CQT-type query (Lykken, 1991). We will use the term polygraph for any combination of psychophysiological recording and a formal deception-generating paradigm.

Brain fMRI is performed by placing the head or the entire body in the bore of an MRI scanner, built around a powerful (× 10⁵ Earth’s magnetic field) electrical magnet. The scanner generates two magnetic fields: the main, perpendicular to the plane of the central bore, and the weaker ‘gradient’ fields, at an angle to the main. The rapid on/off switching of the gradients causes the hydrogen nuclei in the body water to resonate and emit radiofrequency signals that are eventually reconstructed into a three-dimensional image that reflects the relative concentrations of the hydrogen nuclei in the tissues. MRI’s temporal (i.e. the ability to distinguish two points in time) and spatial (i.e. the ability to distinguish between two points in space) resolution can be tailored to the task at hand; a combination of high spatial and low temporal resolution (i.e. sharp images taking a long time to record) is used when brain structure is of interest, while the opposite combination (fast but fuzzy) is used to characterize the transient changes in blood flow and oxygenation accompanying brain’s function. A typical fMRI sequence
has a spatial resolution of 1–3 mm and temporal resolution of 2–3 seconds. There is not a qualitative difference between structural and functional MRI and most modern clinical MRI scanners with main magnetic field strengths above 1.5 Tesla are capable of fMRI (Glover, 2003).

The brain’s electrical activity is of the order of milliseconds, while the resultant changes in the regional brain blood flow are of the order of seconds. The (neurovascular) coupling between neuronal (electrical) activity, metabolism, and perfusion is robust, though it could be modulated by disease or other factors (Hamzei, Knab, Weiller, & Röther, 2003; Logothetis & Pfeuffer, 2004). When local oxygen demand rises in response to increased electrical activity and metabolism, the inflow of oxygenated haemoglobin causes the fMRI signal emanating from the tissues adjacent to and supplied by a particular blood vessel to increase. This is the basis of BOLD (blood oxygenation level-dependent) functional Magnetic Resonance Imaging (fMRI), a common fMRI approach (Ogawa, Menon, Kim, & Ugurbil, 1998). Since there are no normative BOLD fMRI signal values, BOLD fMRI studies usually rely on the difference in brain response between the target and control stimuli. This so-called principle of ‘cognitive subtraction’ makes the selection of appropriate baselines and comparisons critical (Aguirre & D’Esposito, 1999). In summary, fMRI is based on several valid physiological and methodological assumptions, which are inherently vulnerable to the Scylla and Charybdis of translational research: design flaws and individual variability. fMRI has the theoretical advantage over the polygraph of measuring central rather than peripheral nervous system activity and the advantage of spatial resolution - over the EEG. In addition, the nature of the fMRI procedure precludes the interpersonal expectancy effects (Rosenthal, 1994) while the automated and protocolized data processing boosts the reproducibility and objectivity of the results (Aguirre & D’Esposito, 1999; Beckmann, Jenkinson, & Smith, 2003; Phan et al., 2004; Smith et al., 2004).

The reliability, safety, and availability of fMRI have led to attempts to utilize it for lie detection. The body of work correlating brain activity with basic cognitive functions such as attention, response inhibition, and cognitive conflict (Carlson, Moses, & Hix, 1998; Konishi et al., 1999; MacDonald, Cohen, Stenger, & Carter, 2000) provided the basis for the interpretation of the early fMRI deception experiments in terms of basic cognitive operations (Langleben et al., 2002; Spence, 2004). Building on the pioneering electrophysiological (EEG) studies of deception, this body of research set the stage for the initial set of experiments that have demonstrated the feasibility of discrimination between lie and truth under the conditions of forced choice and sometimes even cued response (Langleben et al., 2002; Lee et al., 2002; Spence et al., 2001). Subsequent work diverged into two arms, the first devoted to further refinement of lie-detection techniques and the second to the functional neuroanatomy and cognition of deception (Davatzikos et al., 2005; Ganis, Kosslyn, Stlose, Thompson, & Yurgelun-Todd, 2003; Kozel et al., 2004, 2005; Nunez, Casey, Egner, Hare, & Hirsch, 2005). While the former advanced from group-average to single-participant studies (Kozel et al., 2005; Langleben et al., 2005), the latter demonstrated the potential contributions of memory, learning, and emotion to the brain activity patterns of deception (Abe, Suzuki, Mori, Itoh, & Fuji, 2007; Ganis et al., 2003; Nunez et al., 2005). Both directions converged on a model of deception as a working memory-intensive task, mediated to a large extent by the prefronto-parietal systems dedicated to behavioural control and attention (Langleben et al., 2004; Spence, 2004; Nunez et al., 2005).
The critical issue *en route* from abstract laboratory experiments to a diagnostic test is that of reproducibility, i.e. inter-rater reliability, which in-turn is a function of the accuracy of the measurement technique, the robustness of the activation task, and the individual variability. The technical fluctuations in the accuracy of BOLD fMRI are negligible relative to the range of individual variability in brain structure and function. Replication, a scarce commodity in academic fMRI research, is modestly encouraging in the deception studies. For example, the GKT2 task activation pattern we reported (Langleben *et al.*, 2005) has been independently replicated (Gamer, Bauermann, Stoeter, & Vossel, 2007). More importantly, the principle of increased prefronto-parietal activation during lie has been observed in most fMRI studies of deception across experimental paradigms (Kozel *et al.*, 2005; Lee *et al.*, 2005), as long as countermeasures were not present (Langleben *et al.*, 2005). The initial assumption that the prefronto-parietal activation is lie specific independently of task structure has been refuted by the demonstration that most of this activation can be shifted to the truthful responses through manipulation of the non-specific item salience (Langleben *et al.*, 2004, 2005). The lack of specificity of prefronto-parietal activation to lie has theoretical and applied implications. On the theoretical side, the model postulating that suppression of the prepotent truth is the only basic mechanism of lie, needs to be revised. Traditionally, response inhibition has been associated with right inferior frontal gyrus activation (Konishi *et al.*, 1999). In our study (Langleben *et al.*, 2005), the brain region that has remained impervious to the salience shifts was the left inferior frontal cortex (IFG), encompassing Brodmann areas 45, 46, and 47. This area has been flagged in the early literature on malingering (Ward, Oakley, Frackowiak, & Halligan, 2003) and is the premier locus of behavioural control (Aron, 2007). Thus, deception is likely to engage executive processes other than mere response inhibition. The limbic system has been notably absent from the current data on deception, often by design (Langleben *et al.*, 2002). Recent experiments began uncovering its potential role, making it an additional candidate for inclusion in the revised model of deception (Abe *et al.*, 2007; Hakun, Ruparel *et al.*, in press; Langleben, Dattilio, & Guthiei, 2006). On the applied side, reversal of the expected activation pattern by saliency shifts did not preclude discrimination between lie and truth, because it was based on the absolute value of the differences between lie and truth (Davatzikos *et al.*, 2005; Langleben *et al.*, 2005).

The accuracy of fMRI-based lie detection is another critical translational issue stemming in part from the fundamental problem of individual variability. The estimates from the two available laboratory datasets, ranging between 76 and over 90% (Davatzikos *et al.*, 2005; Kozel *et al.*, 2005; Langleben *et al.*, 2005), should be considered a strong indication for more extensive testing rather than a focus of debate on whether the upper limits of this range is sufficient for court evidence. In practice, the accuracy of fMRI-based lie detection is likely to vary with questionnaire-type, countermeasures, and other, hitherto unexplored variables. Though individual variability may diminish the value of a standard fMRI template in clinical lie detection, individual response patterns to a standard task could be used as a within-participant control prior to every lie-detection test (Hakun, Ruparel *et al.*, in press).

The relatively high degree of reproducibility and accuracy of fMRI-based lie detection in the laboratory has raised academic and public curiosity, commercial activity, and political concerns (Marantz Henig, 2006; Talbot, 2007). The more fervent critics of the new technology have been concerned that it might be effective enough to endanger ‘cognitive freedoms’ (Boire, 2005). This line of criticism fails to make the important distinction between mind-reading and lie-detection applications of fMRI.
Deception is not a thought but a voluntary act (Kant, 1964; St. Augustine, 1948). Thus, while unauthorized access to one’s thoughts and opinions (i.e., mind-reading (Norman, Polyn, Detre, & Haxby, 2006)) may be obtained using fMRI tools developed to study addiction and schizophrenia (Achim, Bertrand, Montoya, Malla, & Lepage, 2007; Lingford-Hughes, 2005), lie detection demands that there is a lie, i.e., a conscious and voluntary response to a query (Langleben et al., 2006). Since stability and accuracy of human memory are limited (Slotnick & Schacter, 2004), mind-reading is a complex endeavor that remains in the domain of basic research (Cox, Meyers, & Sinha, 2004). fMRI-based prediction of behavior, specifically, detection of intent, is also the topic of basic research (Haynes et al., 2007) and science fiction (i.e., ‘Eternal Sunshine of the Spotless Mind’, Focus Features Inc, 2004). Of particular relevance to fMRI-based lie detection is the notion that progress in neuroscience may redefine the legal definition of deception, i.e., fMRI evidence of intent may become a biological criterion for punishable deception (Langleben et al., 2006). At the present time, lie detection is the only fMRI application developed enough to be considered for translation into forensic practice.

The privacy concerns about fMRI that may be justified for mind-reading are strongly mitigated when applied to lie detection because of the conscious nature of deception and the atheoretical ‘black box’ fMRI approach to lie–truth discrimination. Another, often missed ‘safeguard’ against nonconsensual use is the need for the participant’s physical cooperation in studies in which the participant needs to remain almost completely immobile for at least 10 minutes to produce usable fMRI data. Involuntary restraint would be counterproductive, as the stress it would induce may override the subtle fMRI signal differences between lie and truth. Further, though involuntary fMRI could conceivably be legal (Thompson, 2005), it would always violate the medical code of conduct, highlighting the advantage of categorizing forensic fMRI as a medical procedure.

Thus, in the foreseeable future, ethical and legal use of fMRI-based lie detection is likely to be limited to cooperative volunteers, such as those seeking acquittal, or clearance after failing a polygraph test. Different applications may require different accuracy trade-offs. For example, false positive rates not acceptable for incriminating evidence could be sufficient to demonstrate reasonable doubt (Halber, 2007).

One human rights concern stands: the risk of abuse through invalid application or misinterpretation of the data (Wolpe, Foster, & Langleben, 2005). This problem is part of a more general issue of the legitimacy of using a medical device for non-medical purpose. Non-medical use of medical equipment outside research labs would require a new set of ethical, legal, and professional guidelines. Bringing forensic applications of fMRI under the umbrella of comprehensive professional codes guiding physicians and clinical psychologists may substantially reduce the risk of unethical use (Moberg & Kniele, 2006; Rosen & Gur, 2002). Yet another, rather circular criticism posits that it is impossible to conceive of ethically sound prospective clinical trials of fMRI-based lie detection and equally impossible to proceed to real-life cases without such data (Nancy Kanwisher quoted in (Halber, 2007)). This lively debate can be continued ad infinitum as long as the critical translational data is missing.

Designing clinical trials of fMRI-based lie detection implies using unproven technology in situations where balancing the risk/benefit ratio may be legally (Appelbaum, 2007) and ethically challenging (Thompson, 2005) but is by no means impossible (Hakun, Ruparel et al., in press). The results of these trials would determine the true potential and limitations of fMRI-based lie detection and have high societal
impact. Therefore, at least some of them should be funded through peer-reviewed public organizations.

In conclusion, the scope of criticism of fMRI-based lie detection runs the gamut from sound to irrational. Naturally, criticism highlights the downside, while omitting the two important advantages: First, the strong and growing demand for objective lie detection is not met by the existing technology. The demand is so strong that the US government continues to use the polygraph extensively, exempting itself from its own ban or non-government pre-employment polygraph screening (OTA, 1983). Even the presumed inadmissibility of the polygraph in courts does not preclude it from influencing the prosecution's decisions to prosecute. Second, fMRI is unquestionably a qualitative leap forward in our ability to correlate brain activity with behaviour and cognition (Logothetis & Pfeuffer, 2004). Thus, although one cannot predict which combination of cognitive probes and fMRI sequences will ultimately become a method of choice in lie detection, truth verification, and/or information gathering, the prevailing demand and technical feasibility are likely to produce a clinical fMRI-based lie detector in the near future.

Acknowledgements

This work has been supported by the grants from the Defense Advanced Research Projects Administration (DARPA), Army Research Office (ARO), Office of Naval Research (ONR), and career development support from the National Institutes on Drug Abuse (NIDA).

References


Received 11 October 2007; revised version received 12 October 2007
Legal and Criminological Psychology

Publishes original research in all areas of psychology and law:
- victimology
- policing and crime detection
- crime prevention
- management of offenders
- mental health and the law
- public attitudes to law
- role of the expert witness
- impact of law on behaviour
- interviewing and eyewitness testimony
- jury decision making
- deception

The journal publishes papers which advance professional and scientific knowledge in the field of legal and criminological psychology defined broadly as the application of psychology to law and interdisciplinary enquiry in legal and psychological fields.

Editor

Professor Aldert Vrij,
Portsmouth University, UK

“ONE OF THE MAIN STRENGTHS OF THE JOURNAL IS ITS BROAD BASE APPEAL AND INTERNATIONAL SCHOLARSHIP AS WELL AS ITS AIM TO PUBLISH INNOVATIVE AND INTERESTING RESEARCH ARTICLES IN FORENSIC PSYCHOLOGY”

ISSN 1355-3259  Volume 13 (2008)  2 Issues
Non-Member  US$74
Institutional  US$220

Submit articles online @ http://lcp.edmgr.com
View online articles and free sample copies @ www.ingentaconnect.com/content/bpsoc/lcp

www.bpsjournals.co.uk/lcp