University of Pennsylvania

ScholarlyCommons

Neuroethics Publications

Center for Neuroscience & Society

2014

Functional MRI-Based Lie Detection: Scientific and Societal Challenges

Martha J. Farah
University of Pennsylvania, mfarah@psych.upenn.edu

J. Benjamin Hutchinson

Elizabeth A. Phelps

Anthony D. Wagner

Follow this and additional works at: https://repository.upenn.edu/neuroethics_pubs

Part of the Bioethics and Medical Ethics Commons, Criminology Commons, Criminology and Criminal Justice Commons, Neuroscience and Neurobiology Commons, and the Neurosciences Commons

Recommended Citation

Functional MRI-Based Lie Detection: Scientific and Societal Challenges

Abstract

Functional MRI (fMRI)-based lie detection has been marketed as a tool for enhancing personnel selection, strengthening national security and protecting personal reputations, and at least three US courts have been asked to admit the results of lie detection scans as evidence during trials. How well does fMRI-based lie detection perform, and how should the courts, and society more generally, respond? Here, we address various questions — some of which are based on a meta-analysis of published studies — concerning the scientific state of the art in fMRI-based lie detection and its legal status, and discuss broader ethical and societal implications. We close with three general policy recommendations.

Keywords

Disciplines

Bioethics and Medical Ethics | Criminology | Criminology and Criminal Justice | Neuroscience and Neurobiology | Neurosciences

Comments

Correction for the published paper, 19 Feb. 2014:

An incorrect paper was cited as reference 2 of this article. The correct paper is Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A. & Schendan, H. E. Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage* **55**, 312–319 (2011). This has been corrected in the online version.

Farah, M. J., Hutchinson, J. B., Phelps, E. A., & Wagner, A. D. (2014). Functional MRI-based

lie detection: scientific and societal challenges. *Nature Reviews Neuroscience*, *15*(2), 123-131.

fMRI Lie Detection: Scientific and Societal Challenges

Martha J. Farah, University of Pennsylvania

Ben Hutchinson, Princeton University

Elizabeth A. Phelps, New York University

Anthony D. Wagner, Stanford University

Correspondence to:

mfarah@neuroethics.upenn.edu

Abstract

FMRI lie detection has been marketed as a tool for enhancing personnel selection, strengthening national security and protecting personal reputations, and at least three US courts have been asked to admit the results of lie detection scans as evidence at trial. How well does fMRI lie detection perform, and how should the courts, and society more generally, respond? Here we address a sequence of questions concerning the scientific state of the art in fMRI lie detection followed by consideration of its legal status and broader ethical and societal implications. Scientifically, we begin by reporting a meta-analysis of the literature, which indicates that reliable patterns of activation are associated with deception across different tasks and laboratories. However, we also find that other psychological processes are confounded with deception in many tasks and may be responsible for the observed fMRI effects attributed to deception. Finally, we review a number of unmet scientific challenges concerning the generalizability of current laboratory research on lie detection to the real world. Regarding the impact of fMRI on society, we review its current legal status, the standards by which it will continue to be judged, and consider the potential consequences and ethical principles at stake when and if fMRI lie detection were to be adopted in specific contexts. We close with three general policy recommendations.

Over the centuries human beings have devised many different methods for the detection of deception (BOX 1). Some are low-tech, for example the skilled recognition of facial expressions, and some are high, including the polygraph, a device that measures autonomic arousal and is known in popular culture as the "lie detector." At the present time, no method of lie detection has been proven to perform with high accuracy in the field, and the search for a better method continues.[1]

Recent efforts to detect lies have focused on measures of brain function rather than autonomic nervous system activity. The appeal of this approach is that, in contrast to previous methods that detected the emotional arousal resulting from deception, brain-based lie detection measures physiological changes related to cognition during deception, and could therefore, in principle, be detecting the process of deception itself. Most functional imaging attempts to discriminate lying from truth-telling have employed functional magnetic resonance imaging (fMRI), although a few early studies used positron emission tomography (PET) and other methods (event-related potentials and functional near infrared spectroscopy) have been applied to the related problem of detecting concealed knowledge.[2,3] Scientific and legal interest in fMRI lie detection has developed rapidly. The majority of scientific articles have been published within the last decade, and there have been at least three attempts to have fMRI lie detection admitted into United States courts since 2010.

In the present article, we assess the current state of the science in fMRI lie detection and review some of the legal and societal issues raised by this technology. Beginning with the science, we address three questions about the current state of the art in fMRI lie detection. First, do current findings on lie detection, from different laboratories and using different experimental tasks, identify a consistent set of brain regions and, if so, which areas are they? Second, how

confidently can we interpret the results of these studies with respect to the neural substrates of deception per se, and what alternative interpretations have yet to be ruled out? Third, what additional challenges do we face in the effort to use fMRI for the detection of deception in real-world contexts? We then raise a series of issues concerning the ethical, legal and societal impact of attempting to detect lies with fMRI.

The science of fMRI lie detection

Although fMRI lie detection has been commercialized and is used by some for real-world applications, research on this topic began as a form of basic science, undertaken with the goal of identifying the neural systems involved in deception.[4-6] In such studies, blood-oxygen-level dependent (BOLD) activity is measured under conditions in which subjects are instructed or explicitly permitted to make deceptive versus truthful responses. Deception has been operationalized in many different ways in fMRI lie detection research, as in lie detection research more generally (Box 2). The designs of these studies are critical for understanding the degree to which they successfully isolate the neural correlates of deception, so several examples of research tasks will be given here.

In one of the earliest studies, subjects were given two playing cards and were instructed to deny possession of one and acknowledge possession of the other.[5] Subjects were scanned while they viewed a series of cards, including the two critical cards and other cards they had not been given. Activity associated with deception was isolated by comparing the lie and truth trials. A similar task involved having subjects pick a number between 3 and 8, and then deny having picked that number (the critical lie item) and deny having picked the other numbers (truth items) while in the scanner viewing numbers.[7] Here again, activity on lie trials was compared to that on

truth trials to discover which areas were associated with deception in this task.

A small step toward more realistic experimental paradigms to study lying was taken by Andrew Kozel and colleagues.[8] They devised a mock crime scenario, in which subjects were given a choice whether to "steal" a ring or a watch and were instructed to place the chosen item in a locker. Subsequently, they answered questions about the "crime" while in the scanner, and were instructed to deny having taken either item. In addition to lies (denials regarding the item they took) and truthful statements (denials regarding the item they did not take), subjects were asked yes/no general knowledge questions such as "Is it 2004?" or "Do you live in the United States?" The difference between the lie and truth trials was taken to index the neural activity associated with lying. Finally, researchers have compared subjects' responses when answering questions about past events or personal information, for which they were cued to respond truthfully or deceptively.[9]

In sum, for most fMRI studies of lie detection subjects are instructed by the experimenters to lie and to tell the truth on specific trials of the experiment, and the activation from these trials is either directly contrasted or compared after contrasting to a baseline condition. The regions showing significantly greater activation for lies than truth are taken to be the neural correlates of deception.

Consistency of results across laboratories and tasks. When the deceptive and truthful response conditions of tasks are compared, activation is often found in certain regions, particularly prefrontal cortex, anterior cingulate cortex, and parietal cortex. How reliably are these regions associated with deception, and can one identify more specific subregions within these relatively broad anatomical areas that are activated during deception? While earlier reviews

of the literature found substantial consistency across studies [10-12], the literature has continued to rapidly grow in recent years. Accordingly, to assess the consistency of deception-related activity across laboratories and tasks, we carried out a new meta-analysis of the fMRI lie detection literature to date.

Like the earlier analyses of Christ et al.[11] and Wagner[12], the present meta-analysis used the activation likelihood estimation (ALE) method[13] (for details, see **Supplementary Materials**). ALE quantifies the degree of anatomical overlap across published neuroimaging studies based on peak-voxel coordinate information. This enabled us to quantify the reliability of anatomical overlap of activation foci observed, rather than simply analyzing commonalties in terms of activations occurring in pre-selected regions of interest. Published lie detection studies were included if they (1) used fMRI; (2) reported results from a whole-brain, group analysis of healthy, young adults; (3) conducted a statistical contrast indexing deception, reporting one or more foci in standardized coordinate space; and (4) reported data that were not reported (in part or in full) in any other study included in the meta-analysis (that is, a given data set contributed to the analysis only once). Critically, activations analyzed were group-level contrasts of deceptive versus truthful responding (**Table S1**). As detailed in the **Supplemental Materials**, this analysis was performed over 321 foci from 28 independent statistical contrasts between lie versus truth conditions that were reported in 23 different studies.

As shown in **Figure 1** (see also **Table S2**), the meta-analysis revealed a number of regions that were active across studies at an above-chance rate, including bilateral dorsolateral and ventrolateral prefrontal cortex, inferior parietal lobule, anterior insula, and medial superior frontal cortex. As previously noted by others, there was also considerable variability from study to study, as no region was active in all (or nearly all) studies. This may be due in part to

differences in tasks and stimuli[9], in data acquisition procedures (e.g., magnet field strength, acquired functional data resolution, as well as the number of trials per condition and number of subjects, both of which impact statistical power), and in statistical procedures (e.g., choice of statistical thresholds). Nevertheless, across studies there is considerable agreement on the question of which brain areas are more active during instructed lying compared with truth-telling.

Could the patterns documented by this meta-analysis provide the scientific foundation for a useful lie detector? We suggest that several other empirical questions need to be addressed before fMRI lie detection can be considered for real-world use. Are the observed brain activations due to deception per se or to confounds within the experimental designs? More generally, is the observed activation specific to lying or does it reflect something about the way lies are usually (but not necessarily or invariably) produced? Can fMRI discriminate lies from truth in individual subjects with sufficient accuracy to be useful in at least some circumstances? For what types of subjects? Do laboratory-derived indicators generalize to the real world, where stakes and hence emotions may be high, where base rates may be unknown, and where subjects may attempt countermeasures?

Experimental confounds and other questions of specificity. The experimental designs used in fMRI studies of deception are most naturally described in terms such as "lie" and "truth," and this language encourages a certain presumption of specificity. However, on the basis of simple experimental contrasts between lie and truth we cannot know what psychological processes other than deception might evoke the same patterns of activity. What can we conclude from the current literature concerning the specificity of deception-related activation?

In most of the tasks used to study deception with fMRI, a number of experimental factors are confounded with the lie versus truth manipulation.[7, 12, 14, 15] For example, in one study [16, 17], BOLD activity differed between instructed lie and truth trials. However, the frequency of the motor response required on truth trials was much lower than that of the motor response required on lie trials. Because of this confound, rather than reflecting neural correlates of deception, critical aspects of these data may reflect neural responses associated with selecting a frequent versus an infrequent motor action.

In an experiment with even broader implications for the interpretation of extant fMRI lie detection data, Hakun and colleagues[7] carried out an experiment in which subjects were scanned while viewing a series of numbers after having chosen one in advance. In one condition, subjects were instructed to lie, by responding that they had not chosen the number when it was shown during scanning. In a different condition, subjects simply viewed the numbers without responding during scanning. Strikingly, in three out of three subjects, Hakun observed greater lateral prefrontal and parietal activation to the chosen number relative to other numbers when subjects simply passively viewed the numbers as well as when they were instructed to lie about the chosen number. This finding suggests that in many of the studies in our meta-analysis, the greater activity observed for instructed lie stimuli may not reflect neural processes related to deception, but rather may reflect other cognitive differences produced by the task. For example, the mere act of selecting a stimulus at the outset of the experiment (be it a selecting a specific number from the range 3 to 8, or selecting a ring rather than a watch from a drawer) may attach particular significance to the stimulus that alters subsequent cognitive responses to the stimulus when it appears during fMRI scanning. The selected stimulus may be more salient relative to the other ("truth") stimuli, resulting in differential engagement of neural mechanisms of attentional

orienting, and the selected stimulus may also be associated with stronger or richer memories, resulting in differential engagement of neural mechanisms of memory. Are the activation differences between "lie" and "truth" stimuli due to the act of deception, or to such confounding effects of attention and memory?

The presence of memory confounds in fMRI lie detection studies was directly addressed in an important study by Gamer et al.[14] In that study, subjects were instructed to encode into memory critical items (a banknote and playing card), and were then later scanned while simply viewing the critical items and control items (subjects pressed a button to indicate each stimulus's presentation). Because this stimulus viewing task did not require subjects to respond deceptively or truthfully, differences in BOLD signal between the critical items and control items must be due to differences in the items' histories; that is, whether the subject does or does not have a memory for the item. Strikingly, the results revealed greater activation to the critical items versus control items in prefrontal and anterior insular cortical areas previously observed in fMRI studies comparing instructed lie versus truth trials. Because memory and attention confounds are present in many fMRI lie detection studies, these data, along with those of Hakun et al.[7], cast doubt on whether the existing literature bears on the question of whether there are consistent neural correlates of deception per se.

Related to the preceding concerns about experimental confounds is the broader issue of process (or functional) specificity. Even if it were possible to correct the memory and attention confounds inherent in the tasks just described, there would likely remain an association between deception, on the one hand, and executive function and other cognitive processes, on the other, because deception generally places more demands on executive functions and memory compared to truthful responding. The liar must generally keep two versions of events in working memory

and inhibit the more natural response of responding in accordance with reality. It has been noted that the regions associated with deception in the fMRI lie detection literature are also associated with executive function, attention, and memory processes, consistent with this association between deception and cognitive load.[10,11] As many have argued, it is possible that fMRI lie detection is measuring differences in engagement of these more general-purpose cognitive processes. To the extent that deception does not necessarily (but only typically) impose a higher cognitive load than truth telling, it may be possible to dissociate the two. Under certain circumstances, true responding could tax these processes to an equivalent or greater extent than deceptive responding, a pattern which has in fact been found in at least once instance.[18] Thus, despite the encouraging consistency revealed by the meta-analysis (**Figure 1**; **Table S2**), truth-telling could be mistakenly interpreted as deception according to current methods of fMRI lie detection.

Translational challenges: From the laboratory to the real world.

Even if all of the uncertainties just described could be resolved, many other issues would have to be addressed before fMRI lie detection could be used responsibly in the real world. We turn to these issues here.

Inferences about individual subjects. Real-world uses of lie detection will of course involve inferences about the truthfulness or deceit of individuals. What do the published studies tell us about the accuracy with which deception can be identified at the individual level? Most publications report only group analyses, making them poorly suited to answering this question. Of the minority of studies reporting statistics that are directly relevant to assessing accuracy at

the level of individual subjects or individual events [2, 7, 8, 16, 17, 19-27] , only two studies (to our knowledge) report data relevant to detecting deception at the individual event level. Specifically, using the same dataset, Langleben et al.[16] and Davatzikos et al.[17] focused on whether instructed lie and truth events could be discriminated, using either logistic regression16 or non-linear machine learning analyses[17]. Although these event-level analyses yielded accuracy rates of 78% and 88%, respectively, the impact of these early findings is limited because of the above-noted response frequency confound in the experimental design.

Other studies focused on examining whether fMRI BOLD activity differs when an individual is lying compared to telling the truth (pooling data across events). A variety of statistical approaches have been implemented, including single-subject univariate analyses [7,19,20,21], univiariate analyses combined with the counting of above-threshold voxels in targeted regions of interest [8, 22, 25], and machine learning classification.[2] The reported accuracies in these individual-subject level analyses have ranged from 69-100%, suggesting promise. However, here again the noted concerns about attention and memory confounds undermine data interpretation.[7,12,26] The problem of confounds remains whether the neural correlates of deception are sought using univariate or multivariate analyses, and whether the correlates are discovered by simple regression analysis or machine learning algorithms, so long as the data comes from tasks that do not separate the effects of deception from the effects of attention and memory. Furthermore, even high accuracy rates may decline precipitously when subjects use countermeasures in an attempt to conceal their 'deception.' [2]

The laboratory studies assessing the accuracy of fMRI lie detection on the individual-subject level assess the sensitivity and specificity within an individual by differentiating trials on which the individual is deceptive or truthful. However, determining the accuracy of a test in a

general population also requires an assessment of the test's sensitivity and specificity across individuals within that population. That is, what is the likelihood of detecting deception when it is present in a member of the population, and what is the likelihood of correctly indicating when deception is absent. In addition, determining the real-world accuracy of a detection test depends on a critical third factor, which is the probability of the event occurring within the population, or the base rate. To date, fMRI lie detection tests examining accuracy within individuals have generally not assessed the specificity of the test across individuals. One exception is a study by Kozel and colleagues[22], which tested participants that had been successfully classified as 'lying' on a prior mock crime task (25 of 36 participants). These preselected participants were then examined on a secondary mock crime task. On this secondary task, some of the participants committed the mock crime and others did not, but all were instructed to indicate that they didn't. Kozel et al. were able to correctly detect deception in 100% of the participants in the 'mock crime present' condition. However, they also mistakenly detected deception in 67% of the participants in the 'mock crime absent' condition. In the language of diagnostic testing, the sensitivity of this test was high but the specificity was low.

Note that the risks associated with such low specificity will depend on the base rate of lying in the population assessed [29, 30.] Imagine this test were given to 101 people, 100 of them truthful and one deceptive. Based on the false positive rates of Kozel and colleagues the test would identify 68 participants as 'lying' -- the 1 participant who lied about the mock crime and 67 who did not. In other words, given a positive result, the probability of the test accurately indicating someone as lying is 1 in 68, or less than 1.5%, and the likelihood of incorrectly indicating deception when it is not present is over 98%. As this example illustrates, even in an ideal circumstance in which a laboratory lie detection test is developed and used in identical

situations, and is able to detect deception within an individual 100% of the time, its accuracy in a larger population may still be unacceptably low if the specificity of the test is low and the base rate of lying is low.

The real-world validity of fMRI lie detection will also depend on the generalizability of the findings obtained with laboratory subjects (typically healthy, educated young adults, i.e., undergraduates) to the individuals whose veracity is to be assessed by these methods. Consider the differences between criminal offenders, a group likely to be subjected to lie detection methods, and the university samples on which these methods have so far been tested. A relatively high proportion of criminals meet the criteria for psychopathy, a condition associated with frequent acts of deception, which has been linked to differences in both structural and functional MRI studies.[27] A study of fMRI lie detection in criminal offenders with a diagnosis related to psychopathy, specifically anti-social personality disorder, found that a large proportion of these participants did not show typical prefrontal BOLD response patterns during instructed deception.[28]

Cognitive, personality and brain factors associated with a wide range of individual differences may also affect the validity of fMRI lie detection. Structural and functional MRI changes observed with advancing age, a range of psychopathologies (e.g., schizophrenia, PTSD) or individual traits (e.g., high anxiety, extraversion) limit the applicability of lie detection tests that have not been validated in these populations. There may also be important individual differences in the neural systems of deception per se that we have yet to characterize. In a clever study in which participants were given the opportunity to gain money by being dishonest, those who tended toward dishonesty showed increased BOLD signal in regions related to cognitive control when behaving dishonestly *and* when behaving honestly, whereas participants who

tended to be honest did not show this pattern.[18]

Differences between lies in the lab and in real life. Another potential obstacle to real-world lie detection is that the lies examined in laboratory tasks are generally quite unlike those that we would try to detect outside of the laboratory (BOX 2). Although researchers have been concerned with real-world effectiveness, and have presented their studies as "ecologically valid"[29] or "emulat[ing] as closely as possible a real world situation"[30], the tasks differ in many important ways from the situations in which lie detection would be used in the real world. In the laboratory studies, subjects lie because they are instructed to, about matters with little personal relevance, in highly constrained and contrived situations. In addition, the familiarity of the information being concealed and the level of emotion associated with it are typically much lower in laboratory studies than in real life.

Consider the situation in which a lie is highly rehearsed and thus familiar, and the truth is abhorrent. In this circumstance it seems very possible that truth telling is more effortful than lying. There is evidence from both fMRI and behavioral studies suggesting practice or rehearsal may alter the neural signature of deception. As indicated in **Figure 1**, some of the regions commonly activated in fMRI studies of lie detection include prefrontal regions hypothesized to underlie cognitive functions required during the more effortful 'lie' condition. Studies examining practice effects across a number of cognitive tasks routinely show diminished prefrontal cortex activation. [31] This reduction is thought to reflect the diminished executive control required as highly practiced tasks become automatic. An early fMRI study of lie detection found that memorized lies resulted in less BOLD activation when compared to unpracticed lies in every deception-related region of interest identified except for one hypothesized to be involved in

14

memory retrieval.[9] In a behavioral deception study, it was found the training on deception eliminated pre-training reaction time differences between deception and truth trials, consistent with enhanced automaticity of lying.[32] Outside the laboratory, if one is anticipating interrogation about a lie it is very likely the lie will be practiced and memorized, which might eliminate many of the detectable differences in the behavioral and neural expression of deception.

In addition, real-world deception is likely to be highly emotional and personally relevant. Emotion could impact the neural circuitry of lying in two ways that might make it more difficult to distinguish truth from lies. First, truthfully answering questions about highly emotional events may be more effortful or require more (emotional) control/inhibition than truthfully answering questions about neutral events. To the extent that a lie detection test is measuring non-specific brain signals of effort or inhibition, it may be more likely a true statement about an emotional event is classified as a lie (as has been found with event-related potentials by Proverbio et al., 2013).[33] However, it is also the case that emotional qualities of the event may shift the neural signature of deception to make it less likely a lie is detected. In fMRI studies, emotion has been shown to alter the neural circuitry of memory[34], inhibition and cognitive control[35], and working memory interference[36], all processes believed to underlie the differences between brain activity accompanying deceptive and truthful responses. Indeed, emotional valence has been found to affect the neural localizations of deception-related processing.[37] If a lie detection test is developed based on lying about non-emotional events, its applicability to assessing deception concerning emotional or important, personally relevant events may be limited.

Countermeasures. Methods of lie detection inevitably spawn methods designed to evade detection, so the mere possibility of countermeasures should not be grounds for rejection of

fMRI lie detection. However, the ease and success of countermeasures are relevant to the real-world usefulness of fMRI lie detection. In the one study cited earlier reporting 100% accuracy, the investigators further demonstrated that if subjects adopt a simple countermeasure strategy of making imperceptible finger and toe movements, accuracy fell to 33%. At present researchers have only begun to explore possible countermeasures for fMRI lie detection. Unpublished data from the lab of ADW, focused on whether subjects can conceal memory-related patterns of BOLD signal, also indicate that countermeasure strategies can reduce machine learning decoding of memory states from well above chance to chance levels.[38,39]

In sum, even if the challenges facing fMRI studies of deception in the laboratory, described in the previous section, were met, a number of additional challenges await the successful translation of this method into the real world. The accuracy with which individual subjects can be assessed when telling lies or truth, and the suitability of these accuracy rates given the base rates of lying and truth-telling in the population, demands a major new empirical research effort. How these accuracies vary as a function of an individual's age, health, personality, life history and other variables are also crucial questions and would require an even larger program of research to adequately address. Just as differences among individuals would be expected to influence the validity of the method, so too would differences in the nature of the lie and its context: whether the subject has lived with the lie for a long time, whether the truth is more emotionally charged than the lie, and what is at stake if the lie is discovered. Finally, the susceptibility of fMRI lie detection to countermeasures would need to be more fully explored before it is applied in the real world, and this too requires extensive research.

Current status of real-world fMRI lie detection

FMRI lie detection has moved rapidly from the basic research lab into commercial application in the real world. In 2006 two companies began offering fMRI lie detection services, No Lie MRI, based on the method developed by Langleben and colleagues at the University of Pennsylvania, and Cephos, based on the method of Kozel and colleagues at the University of Texas. These companies have suggested a number of uses for fMRI lie detection, spanning business, family life, criminal justice and national security contexts. For example, No Lie MRI recommends its services for such diverse problems as combating insurance fraud (http://www.noliemri.com/customers/GroupOrCorporate.htm), increasing public trust of US and foreign leaders (http://www.noliemri.com/customers/Government.htm) and "risk reduction in dating" (http://www.noliemri.com/customers/Individuals.htm).

The potential application of fMRI lie detection that has received the greatest public scrutiny has been for assessing the truthfulness of legal testimony. In at least three cases, US courts have been asked to admit evidence from fMRI lie detection. The courts determine admissibility by applying one of two standards, depending on jurisdiction. Both standards are designed to keep "junk science" from influencing jury decisions, although they do so in different ways. According to the Frye standard, set forth in Frye v United States in 1923, admissibility hinges on the general acceptance of the method within its particular scientific field. According to the Daubert standard, set forth in 1993 by the US Supreme Court in Daubert v. Merrell Dow, judges in federal cases must be more active gate-keepers of scientific evidence in court, rather than simply deferring to general scientific opinion. In doing so, they should take five factors (among others they may consider relevant) into account in making their decisions: whether the method is testable, and has been tested; whether it has been reported in peer-reviewed publications; whether there is a known or potentially knowable error rate; whether there are

standards for the way in which the method is used; and finally, as in Frye, whether the method is generally accepted within the relevant scientific community.

The first two attempts to introduce fMRI lie detection occurred in 2010. In the first case, Wilson v. Corestaff Services, the plaintiff in an employment discrimination case sought to introduce evidence gathered by Cephos to support the credibility of a witness' testimony. The judge in this case ruled the evidence inadmissible on the grounds that credibility assessment is the job of the jury, but also noted that the method would not meet the Frye criteria, stating "even a cursory review of the scientific literature demonstrates that the plaintiff is unable to establish that the use of the fMRI test to determine truthfulness or deceit is accepted as reliable in the relevant scientific community" (Wilson v. Corestaff Services, 2010).

In the second case, US v Semrau, the defendant standing trial for Medicare fraud claimed that he did not intentionally violate the law and sought to present the results of fMRI lie detection as evidence that his testimony was truthful. A hearing was held in Memphis Federal Court, before a Magistrate Judge to whom had been delegated the task of making a recommendation on admissibility, to determine whether fMRI lie detection meets the Daubert criteria. The 39-page opinion included the recommendation (later accepted by the presiding district judge) that the fMRI evidence be excluded. The analysis noted: (a) the lack of general acceptance for the method within the scientific community, (b) the substantial differences between laboratory research designs and the real-world use of fMRI lie detection in this case, (c) the lack of "real-life" error rates; and d) the lack of suitably controlling standards for the use of the method. (The latter determination was prompted by Cephos' discounting of one of three scanning sessions, which had indicated deception, on the grounds that the defendant was fatigued.) However, the opinion also observes that precise validation in real-world contexts, of

the sort that scientists might in their own research require, is not always legally necessary: "in the future, should fMRI-based lie detection undergo further testing, development, and peer review, improve upon standards controlling the technique's operation, and gain acceptance by the scientific community for use in the real world, this methodology may be found to be admissible even if the error rate is not able to be quantified in a real world setting" (US v Semrau, 2010).

In the case of *Smith v. State of Maryland*, the defendant was being retried for second-degree murder in 2012 and sought to introduce evidence of the truthfulness of his own testimony from No Lie MRI. The judge refused to admit the evidence based on the Frye standard, after concluding that experts in the field (including ADW and EAP) did not agree with the company's experts, suggesting lack of "general acceptance".

Legal, social and ethical considerations

Science can, in principle, tell us the accuracy of fMRI lie detection for any particular population of individuals under any particular circumstances we might specify. It cannot, however, tell us how accurate fMRI lie detection *should* be for any particular use to which it might be put. That decision depends on the needs and values of the people using the method. We have seen that the legal system has criteria, developed in legal cases that involved scientific evidence other than fMRI lie detection, for deciding when a method has sufficient validation, namely the Frye and Daubert criteria. These criteria make clear the interdependence and the independence of scientific and societal decision-making. While scientific research provides essential input into decisions regarding admissibility, the decisions themselves are not made by the kinds of conventional scientific criteria applied, for example, by reviewers of research articles. It is fair

to say that, in some respects, legal standards may be lower than scientific standards where scientific evidence such as fMRI lie detection is concerned. As argued by Fred Shauer in an article entitled "Can Bad Science Be Good Evidence?," the societal needs served by the law often require a more pragmatic approach to the vetting of scientific evidence. In the absence of better methods of discovering the truth, an imperfect method may be better than nothing: "… the exclusion of substandard science, when measured by scientific standards, may have the perverse effect of lowering the accuracy and rigor of legal fact-finding, because the exclusion of flawed science will only increase the importance of the even more flawed non-science that now dominates legal fact-finding."[40] As Langleban and Moriarity have noted[2013], this reasoning can be extended to other scientific methods as well, given that many of the methods of forensic science have been found wanting. [NRC2009]

As scientists we are not accustomed to endorsing methods on the grounds that they are "lesser evils." The difference between the scientific and legal approach to accepting questionable sources of evidence rests in part on the scientist's choice of hypotheses to test. If no good method is available to test a certain hypothesis, then a scientist will normally simply decline to test it. The legal system cannot make the analogous decision; the question of "guilt beyond a reasonable doubt" must be addressed with the evidence available, even when that evidence is acknowledged to have serious weaknesses.

What about societal decision-making regarding potential uses of fMRI lie detection outside the courtroom? The world has only begun to engage with the question of how best to use, and limit the use of, fMRI lie detection. Questions concerning the necessary degrees of accuracy and validity will undoubtedly require different answers for different tasks in different contexts, according to the potential benefits of correct lie detection, the costs of wrong calls, and

the intersection of this technology with moral principles such as the right to privacy.

How accurate is accurate enough? The most immediate ethical and social issues raised by fMRI lie detection arise because of the method's lack of demonstrated accuracy and validity. Given the scientific and technical problems reviewed earlier, the most likely harms would result from false determinations, both lies wrongly identified as truths and truthful statements wrongly identified as lies. The history of polygraphy offers tragic reminders of the cost, in national security and human life, of overreliance on an apparently high tech but inaccurate method for detecting deception. Some commentators have suggested a ban or moratorium on fMRI lie detection, pending better evidence concerning its accuracy.[41,42]

In two well known instances of the polygraph's false negatives, the American CIA agent Aldrich Ames and the Jordanian CIA informant Humam Khalil Abu-Mulal al-Balawi both passed polygraph testing, twice in the case of Ames. Yet Ames spent years selling American secrets to the Soviets and Russians, and al-Balawi killed 7 CIA agents as a double agent. False positive polygraph results have cost honest people job opportunities and even their liberty, as when they lead to false confessions.[43]

Marks has pointed out that the impressive visual appearance of fMRI lie detection results may eclipse concerns about the method's technical weaknesses, including the likelihood of false positives. He suggests that government agents interrogating detainees would naturally tend to increase the aggressiveness of their tactics if a detainee's fMRI indicates deception. Thus, whereas lie detection methods might be thought to reduce the use of harsh interrogation, they might instead be used to justify abusive treatment of detainees, a particularly deplorable outcome in the case of false positive results.[44]

As noted earlier, false positives in fMRI lie detection will have the greatest negative impact when the method is used to identify relatively rare cases of dishonesty in a population. When applied to large numbers of mostly honest people, even modest false positive rates will result in many wrongly accused individuals. Applications such as the routine screening of job applicants or travelers are therefore problematic, as they would likely result in large numbers of falsely accused individuals in relation to the occasional correct identification. As with other questions concerning requirements for accuracy, the question of how many individuals we are willing to falsely accuse of deception for the sake of an occasional correctly identified liar will be determined not by research but by society's needs and values.

Ethical issues beyond validity and accuracy. Not all ethical issues surrounding fMRI lie detection depend on the method's accuracy; even a technically perfect lie detector would raise ethical issues and be subject to societal deliberation and regulation. Indeed, some issues would become more pressing in the event of successful fMRI lie detection.

FMRI lie detection raises privacy issues that require societal control, much as we place limits on other practices that intrude on privacy from DNA collection to wire-tapping. For example, if everyone's phone conversations and email messages were generally available to their families, employers and the state, and if the DNA of all citizens were on file with law enforcement authorities, much crime and misbehavior would be discovered or, better still, averted. Yet societies place limits on the collection and use of such information in order to protect personal privacy. An additional reason to limit access to such information is to increase the benefit it provides to those in possession of its. Societal management of fMRI lie detection would presumably be aimed at balancing the cost to individual privacy against the collective

22

benefits of reduced crime and terrorism, enhanced personnel selection and the generally increased honesty between people that might result from the knowledge that the veracity of one's statements could be tested.

The legal framework for the protection of privacy includes the fourth and fifth amendments, which several legal scholars have discussed in relation to fMRI lie detection and other forms of brain imaging that provide psychological information.[45-47] The fourth amendment protects against warrantless search, including physical tests such as fMRI. The fifth amendment protects against compelled self-incriminating testimony. It remains to be decided whether fMRI lie detection should be viewed as physical evidence or testimony.[48]

Considerations of individual autonomy and freedom arise in connection with the process of consenting to undergo fMRI lie detection. Consent procedures therefore constitute another aspect of societal management that arises regardless of the method's known accuracy. At present, when accuracy in the real world is unknown and even laboratory accuracy estimates are unavailable for subjects of different ages and states of health, subjects must understand the questionable accuracy of the method in order to make an informed decision on their own behalf. They must also understand that not all outside parties will interpret the results of the test with appropriate caution. Even if accuracy is better demonstrated than at present, subjects must be informed about the inevitably imperfect performance of the method. Consent is especially fraught when testing is not requested by the subject but by another party, for example, the government, an employer or a jealous spouse. Additional safeguards may be needed to prevent coercion, including the indirect coercion that results when refusal to take the test is seen as indicative of guilt.

In sum, the question of whether and how to use fMRI lie detection cannot be answered

23

solely on the basis of the method's performance. No method will ever be known to have 100%

accuracy for any context in which it might be deployed. Deciding what level of uncertainty is

acceptable depends on how different kinds of outcomes are valued. Correct and incorrect

identifications of lies and of truth may be weighted very differently under different

circumstances and in different societies. For example, the priority placed on outcomes related to

security relative to the rights of individuals will determine whether it is worse to miss a liar or

falsely accuse an honest person. The strength of a society's commitments to principles including

individual privacy, autonomy and freedom will also shape its policies concerning fMRI lie

detection.

Policy recommendations

How should the development and use of fMRI-based lie detection be managed, in light of

the scientific, legal, ethical and societal issues just reviewed? We offer three general

recommendations.

First, different policies should be considered for different applications of fMRI lie

detection. We do not join calls to ban fMRI lie detection across the board. Despite the enormous

shortcomings of current evidence, reviewed earlier, we suggest that restrictions should be

proportional to the outcomes and principles at stake. Risk reduction in dating calls for different

standards of certainty and different protections of individual rights than the interrogation of

terrorist suspects.

Second, publicly funded research should be undertaken to explore the potential of fMRI

lie detection, paying attention to conflicts of interest for researchers associated with companies

that offer the service. The two highest research priorities are, first, the removal of, or accounting

for, experimental confounds noted earlier and, second, the validation of the methods under more realistic conditions, including with countermeasures and with more diverse subjects. If fMRI lie detection passes these hurdles, then a substantial investment in real-world validation [44] of the kind proposed by Peter Imrey[50] would be justified.

Third, while acknowledging that the standards of science with regard to truth and certainty may not always be the appropriate ones in legal and other societal contexts, scientists have a vital role to play in the application of neuroscience to the law[51]. In the case of fMRI lie detection, it is our duty to raise questions about the accuracy, validity and specificity of fMRI lie detection, to provide accurate answers to these questions, and to communicate their relevance to nonscientist citizens.

Acknowledgements

Figure 1. Results of the ALE analysis of the fMRI "deception" literature. Overlay of map of ALE values (orange) on the lateral (top) and medial (bottom) inflated PALS surface (Van Essen, 2005) revealing regions consistently implicated in deception across studies. Thresholded at p < .05, False Discovery Rate corrected (per the method described in Eickhoff et al. 2009). MFG: Middle frontal gyrus; IFG: Inferior frontal gyrus; IPL: Inferior parietal lobule; m/SFG: medial/superior frontal gyrus
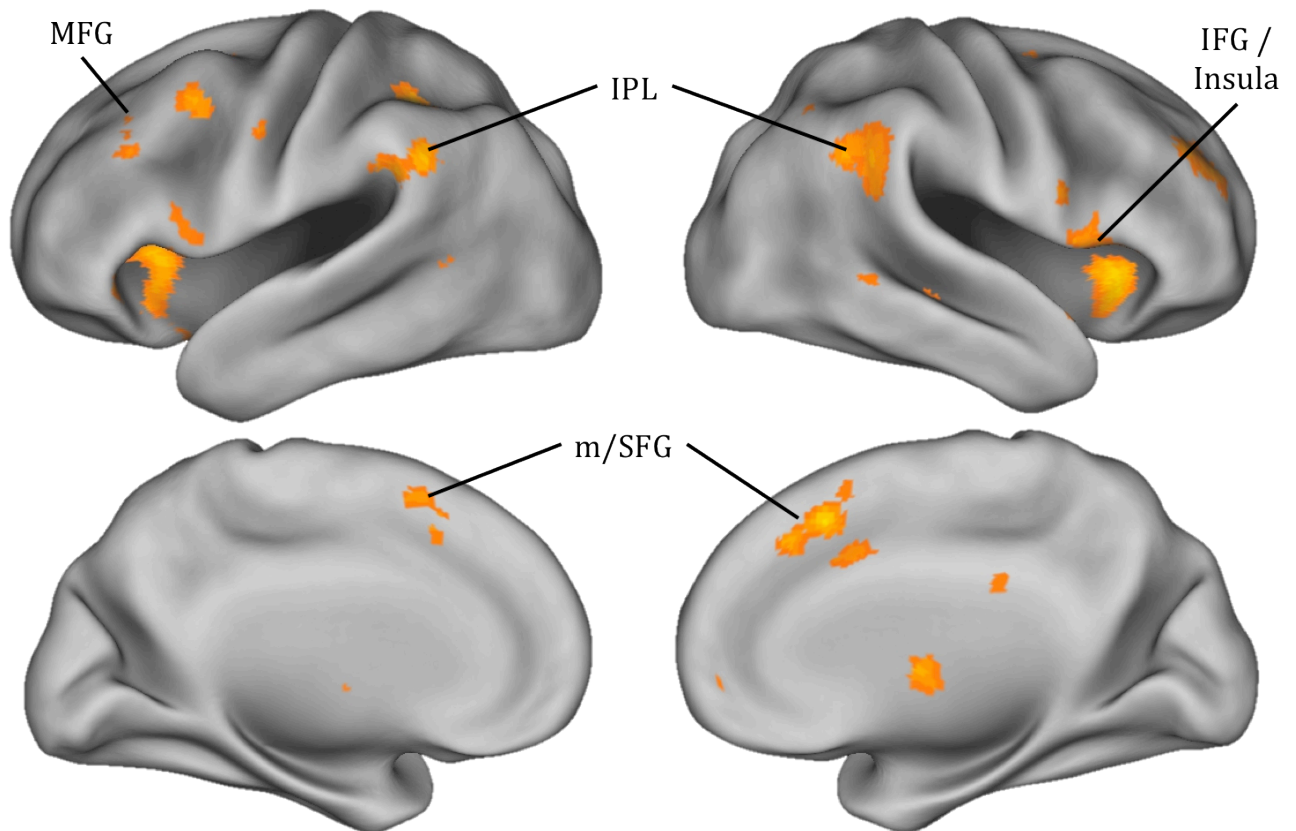
References

1. National Research Council, Committee to Review the Scientific Evidence on the Polygraph. *The Polygraph and Lie Detection*. (National Academies Press, 2003).

2. Ganis, G. & Rosenfeld, J.P. (2011). Neural correlates of deception. In J. Illes and B. J. Sahakian (Eds.), *Oxford Handbook of Neuroethics*. (Oxford University Press, 2011).

3. Heckman, K. E., & Happel, M. D. Mechanical detection of deception: A short review. In: *Educing Information—Interrogation: Science and Art*. Foundations for the Future. Intelligence Science Board, Phase 1 Report, National Defense Intelligence College, (NDIC Press, 2006)

4. Spence, S. A. *et al.* Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* **12,** 2849–2853 (2001).

5. Langleben, D. D. *et al.* Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* **15,** 727–732 (2002).

6. Lee, T. M. C. *et al.* Lie detection by functional magnetic resonance imaging. *Hum Brain Mapp* **15,** 157–164 (2002).

7. Lykken, D. T. Why (some) Americans believe in the lie detector while others believe in the guilty knowledge test. *Integrative Physiological and Behavioral Science*, **26,** 214-222 (1991).

8. Hakun, J. G. *et al.* fMRI investigation of the cognitive structure of the Concealed Information Test. *Neurocase* **14,** 59–67 (2008).

9. Kozel, F. A. *et al.* Detecting deception using functional magnetic resonance imaging. *Biol. Psychiatry* **58,** 605–613 (2005).

10. Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L. & Yurgelun-Todd, D. A. Neural

Correlates of Different Types of Deception: An fMRI Investigation. *Cereb. Cortex* **13,** 830–836 (2003).

11. Bles, M. & Haynes, J.-D. Detecting concealed information using brain-imaging technology. *Neurocase* **14,** 82–92 (2008).

12. Christ, S. E., van Essen, D. C., Watson, J. M., Brubaker, L. E. & McDermott, K. B. The Contributions of Prefrontal Cortex and Executive Control to Deception: Evidence from Activation Likelihood Estimate Meta-analyses. *Cereb. Cortex* **19,** 1557–1566 (2009).

13. Wagner, A. Can neuroscience identify lies? In M. S. Gazzaniga & J. S. Rakoff (Eds.), *A Judge's Guide to Neuroscience: A Concise Introduction*. (University of California, Santa Barbara 2010).

14. Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P.T. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Hum Brain Mapp* **30**, 2907-2926 (2009).

15. Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P. & Vossel, G. fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Soc Cogn Affect Neurosci* **7,** 506–515 (2012).

16. Kanwisher, N. The use of fMRI in lie detection: What has been shown and what has not. In E. Bizzi et al. (eds.), *Using Imaging to Identify Deceit* (pp. 7-13). (American Academy of Arts and Sciences, 2009).

17. Langleben, D. D. *et al.* Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping* **26,** 262–272 (2005).

18. Davatzikos, C. *et al.* Classifying spatial patterns of brain activity with machine learning

methods: application to lie detection. *Neuroimage* **28,** 663–668 (2005).

19. Greene, J. D. & Paxton, J. M. Patterns of neural activity associated with honest and dishonest moral decisions. *PNAS* **106,** 12506–12511 (2009).

20. Kozel, F. A. *et al.* A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *J Neuropsychiatry Clin Neurosci* **16,** 295–305 (2004).

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

22. Monteleone, G. T. *et al.* Detection of deception using fMRI: Better than chance, but well below perfection. *Social Neuroscience* **4,** 528–538 (2009).

23. Kozel, F. A. *et al.* Functional MRI Detection of Deception After Committing a Mock Sabotage Crime*. *Journal of Forensic Sciences* **54,** 220–231 (2009).

24. Jin, B., Strasburger, A., Laken, S. J., Kozel, F. A., Johnson, K. A., George, M. S., & Lu, X. Feature selection for fMRI-based deception detection. *BMC Bioinformatics* **10** Supplement 9, S15 (2009).

25. Nose, I., Murai, J. & Taira, M. Disclosing concealed information on the basis of cortical activations. *NeuroImage* **44,** 1380–1386 (2009).

26. Lee, T. M. C., Leung, M.-K., Lee, T. M. Y., Raine, A. & Chan, C. C. H. I want to lie about not knowing you, but my precuneus refuses to cooperate. *Sci Rep* **3,** 1636 (2013).

27. Ganis, G., Morris, R. R. & Kosslyn, S. M. Neural processes underlying self- and other-related lies: An individual difference approach using fMRI. *Social Neuroscience* **4,** 539–553 (2009).

28. Anderson, N. E. & Kiehl, K. A. The psychopath magnetized: insights from brain

imaging. *Trends in Cognitive Sciences* **16,** 52–60 (2012).

29. Committee to Review the Scientific Evidence on the Polygraph. *The Polygraph and Lie Detection.* (National Academy Press, Washington, 2003).

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

31. Jiang, W. *et al.* A functional MRI study of deception among offenders with antisocial personality disorders. *Neuroscience* **244,** 90–98 (2013).

32. Mohamed, F. B. *et al.* Brain Mapping of Deception and Truth Telling about an Ecologically Valid Situation: Functional MR Imaging and Polygraph Investigation— Initial Experience1. *Radiology* **238,** 679–688 (2006).

33. Abe, N. The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. *Current Opinion in Neurology* **22,** 594–600 (2009).

34. Raichle, M. E. *et al.* Practice-related Changes in Human Brain Functional Anatomy during Nonmotor Learning. *Cereb. Cortex* **4,** 8–26 (1994).

35. Hu, X., Chen, H. & Fu, G. A repeated lie becomes a truth? The effect of intentional control and training on deception. *Front. Psychology* **3,** 488 (2012).

36. Proverbio, A. M., Vanutelli, M. E. & Adorni, R. Can You Catch a Liar? How Negative Emotions Affect Brain Responses when Lying or Telling the Truth. *PLoS ONE* **8,** e59383 (2013).

37. Phelps, E. A. Emotion and cognition: Insights from studies of the human amygdala. *Annu Rev Psychol* **57,** 27–53 (2006).

38. Bush, G., Luu, P. & Posner, M. I. Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences* **4,** 215–222 (2000).

39. Levens, S. M. & Phelps, E. A. Emotion processing effects on interference resolution in working memory. *Emotion* **8,** 267–280 (2008).

40. Lee, T. M. C., Lee, T. M. Y., Raine, A. & Chan, C. C. H. Lying about the Valence of Affective Pictures: An fMRI Study. *PLoS ONE* **5,** e12291 (2010).

41. Uncapher, M. R., Chow, T., Rissman, J., Eberhart, J., & Wagner, A. D. Strategic influences on memory expression: Effects of countermeasures and memory strength on the neural decoding of past experience. *Society for Neuroscience Abstracts* (2012).

42. Rissman, J., Greely, H. T. & Wagner, A. D. Detecting individual memories through the neural decoding of memory states and past experience. *PNAS* **107,** 9849–9854 (2010).

43. Shauer, F. Can Bad Science be Good Evidence - Neuroscience, Lie Detection, and Beyond. *Cornell L. Rev.* **95,** 1191 (2009).

44. Langleben, D. D. & Moriarty, J. C. Using brain imaging for lie detection: Where science, law, and policy collide. *Psychology, Public Policy and Law*, **19**, 222-234 (2005).

45. National Research Council of the National Academies. *Strengthening forensic science in the United States: A path forward*. (Washington, D. C.: National Academies Press, 2009).

46. Greely, H. T. & Illes, J. Neuroscience-Based Lie Detection: The Urgent Need for Regulation. *Am. J.L. & Med.* **33,** 377 (2007).

47. Tovino, S. A. Imaging Body Structure and Mapping Brain Function: A Historical Approach. *Am. J.L. & Med.* **33,** 193 (2007).

48. Costanzo, M. & Krauss, D. *Forensic and Legal Psychology*. (Macmillan, 2010).

49. Marks, J. H. Interrogational neuroimaging in counterterrorism: a 'no-brainer' or a human rights hazard? *Am J Law Med* **33,** 483–500 (2007).

50. Farahany, N. A. Incriminating Thoughts. *Stanford Law Review* **64,** 351 (2012).

51. Pustilnik, A. C. *Neurotechnologies at the Intersection of Criminal Procedure and Constitutional Law*. (Social Science Research Network, 2012). at <http://papers.ssrn.com/abstract=2143187>

52. Shen, F. X. Neuroscience, mental privacy, and the law. *Harvard Journal of Law and Public Policy,* **36,** 653-713 (2013).

53. Stoller, S. E. & Wolpe, P. R. Emerging Neurotechnologies for Lie Detection and the Fifth Amendment. *Am. J.L. & Med.* **33,** 359 (2007).

54. Imrey, P. Validating the neurodeceptograph. Unpublished manuscript, Cleveland Clinic

55. Jones, O. D., Wagner, A. D., Faigman, D. L., & Raichle, M. E. Neuroscientists in court. *Nature Reviews Neuroscience*, **14**, 730-736 (2013).

56. Bond, C. F., & DePaulo, B. M. Accuracy of deception judgments. Personality and Social Psychology Review, **10**, 214-234 (2006).

57. Ekman, P., & Friesen, W. V. *Nonverbal leakage and clues to deception*. (Langley Porter Neuropsychiatric Inst, San Francisco CA Nonverbal Behavior Research Project, 1969).

58. Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. Vocal cues to deception: A comparative channel approach. Journal of psycholinguistic research, **14**, 409-425 (1985).

59. DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. Cues to deception. Psychological Bulletin, **129**, 74-118 (2003).

60. Vrij, A. *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). (Wiley Chichester, UK, 2008).

61. Vrij, A. Nonverbal Dominance Versus Verbal Accuracy in Lie Detection A Plea to Change Police Practice, *Criminal Justice and Behavior*, **35** 1323-1336 (2008)

62. Schafer, E. D. Ancient science and forensics. In Ayn Embar-seddon, Allan D. Pass (eds.). *Forensic Science*. (Salem Press, 2008).

63. Bunn, G.C. The Lie Detector, Wonder Woman and Liberty: The Life and Works of William

Moulton Marston. *Hist Human Sci* **10**: 91–119 (1997).

64. Ben-Shakhar, G., Bar-Hillel, M., & Lieblich, I. Trial by polygraph: Scientific and juridical issues in lie detection. Behavioral Sciences & the Law, **4**, 459-479 (1986).

65. Alder, K. *The lie detectors: The history of an American obsession*. (Simon and Schuster, 2007).

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

67. Aggarwal, N. K. Neuroimaging, culture, and forensic psychiatry. J Am Acad Psychiat Law **37**, 239-244 (2009).

68. Tian, F., Sharma, V., Kozel, F. A., & Liu, H. Functional near-infrared spectroscopy to investigate hemodynamic responses to deception in the prefrontal cortex. Brain research, 1303, 120-130 (2009).

69. Bok, S. *Lying: Moral choice in public and private life*. (Random House Digital, Inc., 2011).

70. Barnes, J.A. *A pack of lies: Towards a sociology of lying*. (Cambridge University Press,1994).

71. DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., & Epstein, J.A. Lying in everyday life. *Journal of Personality and Social Psychology*, **70**, 979 (1996).

72. Ben-Shakhar, G., Bar-Hillel, M., & Kremnitzer, M. Trial by polygraph: Reconsidering the use of the guilty knowledge technique in court. *Law and Human Behavior*, **26**, 527-541 (2002).