Journal of Forensic Radiology and Imaging 8 (2017) 17–21

Contents lists available at ScienceDirect

Journal of Forensic Radiology and Imaging

journal homepage: www.elsevier.com/locate/jofri

Deception detection based on neuroimaging: Better than the polygraph?

Ewout H. Meijer[a,*], Bruno Verschuere[b]

[a] Forensic Psychology Section, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[b] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

A B S T R A C T

Polygraph tests have been used to detect deception for almost a century. Yet for almost as long, the validity of these tests has been criticized. Over the last decade, the use of brain imaging – most notably fMRI - for the detection of deception has attracted increased attention. The expectation is that fMRI can overcome – at least some of - the shortcomings of the polygraph. In this review, we discuss whether this expectation is warranted. Based on our review of the empirical evidence, we argue that fMRI deception research has boosted the theory development of deception. But for practical purposes, fMRI research has thus far done surprisingly little to solve or circumvent the problems than have been associated with deception research for ages.

1. Introduction

The use of brain imaging technology to detect deception has attracted increased attention over the last decade. Take, for example, the following case. In 2003, in the United Kingdom, a 42-year-old woman was convicted of a crime against a child in her care. She served her prison term, yet continued to profess her innocence, even after she was released. Four years after the conviction, psychiatrist Sean Spence administered a deception test based on functional magnetic imaging (fMRI) to assess her credibility. Whilst in the scanner, the woman was presented with statements about the incident (e.g., 'You were innocent of the charges') to which she responded by pressing buttons marked 'yes' or 'no'. Based on the neuroimaging data, Spence and colleagues concluded that her functional anatomical parameters behaved as if she were innocent [49].

Before we evaluate the validity of deception tests such as the one described above, it is worth noting that it is not a coincidence that a medical doctor performed this test. In the earlier days of fMRI research it were the medical specialists – most notably the psychiatrists and radiologists - who had both the interest in deception and the access to fMRI scanners, and they are responsible for much of the early work (e.g., [25,27]). At a later stage, neuroscientists, legal, and ethical scholars became involved in the field. In many of their publications, the 'new' fMRI based deception detection is contrasted with the 'old' polygraph, and it is implicitly or explicitly assumed that fMRI can overcome – some of – the shortcomings of the polygraph (e.g., [6,25,28,11]).

In this contribution, we discuss whether this expectation is warranted: Can fMRI based deception detection help to overcome the shortcomings of the polygraph? We discuss they key difficulties with polygraph testing. We then evaluate to what extent fMRI based deception detection has overcome the problems related to polygraph testing detection. Our analyses will show that fMRI deception research has boosted the theory development of deception. fMRI based deception detection, however, faces largely the same problems as lie detection through the use of a polygraph. For practical purposes, fMRI research has done surprisingly little to solve or circumvent the problems than have been associated with deception research for ages.

2. The polygraph and its questioning formats

The polygraph as we still know it today was first introduced in the 1920's by physiologist and police officer John Larsson from the University of California, Berkley [30]. He developed a machine that simultaneously measured blood pressure, pulse, respiration, and palmar sweating, and used this machine in over a 100 cases to evaluate whether the defendant was telling the truth or not. The polygraphs used today do not fundamentally differ from the one developed by Larsson in the 1920s. The lengthy rolls of paper that the physiological signals were recorded on have been replaced by laptop computers, but the machines still record multiple physiological signals, typically cardiovascular measures, respiration and skin conductance.

The physiological signals measured by the polygraph – or by the fMRI scanner for that matter - can be regarded as an outcome measure. A meaningful interpretation of an outcome measure fully depends on the level of control over the independent variable. For example, to

* Corresponding author.
E-mail address: Eh.meijer@maastrichtuniversity.nl (E.H. Meijer).

http://dx.doi.org/10.1016/j.jofri.2017.03.003
Received 7 November 2016; Accepted 12 March 2017
Available online 15 March 2017

establish whether a form of treatment is effective, a double blind placebo controlled trial is preferred. Only under such controlled circumstances do the treatment and the control groups differ on only one dimension, namely that of the treatment, and can any change in outcome measure be attributed to the treatment. In a similar fashion, the validity of deception detection techniques to a large degree depend on to what extent the questioning format isolates deception. Before one can meaningfully discuss the validity of polygraph or fMRI based deception test, a short evaluation of the most used question formats is crucial. This is why we will shortly explain the three main question formats used in research and/or practice, namely the Control Question Test, the Concealed Information Test, and the Differentiation of Deception test.

The question format most widely used by law enforcement agencies worldwide is the Control Question Test (CQT; [45]). In this type of test, the suspect answers relevant and control questions whilst physiological reactions are being recorded. The relevant questions refer specifically to the incident under investigation (e.g., "In the night of Nov, 3, did you stab X?"). The responses to this question are compared to those elicited by the control questions. These control questions have a more generic nature, but also deal with undesirable behavior (e.g., "In the first 25 years of your life, have you ever done anything illegal?"). The rationale behind the CQT is that for guilty suspects the relevant questions will pose the biggest threat, and will therefore elicit the strongest physiological responses. An innocent suspect, on the other hand, is thought to perceive the control questions as most threatening, and these questions will therefore elicit the strongest physiological responses [42].

Although the use of the CQT is widespread, its merits have been debated for decades. A full review of this debate is outside the scope of the current manuscript, and can be found elsewhere (e.g., [2,38,39]). We focus here on the main criticisms voiced against the CQT. At the core of the debate surrounding the CQT is the general assumption that the relevant questions will elicit stronger emotions – and thus larger responses – only in guilty suspects. Critics argue that this assumption has no basis in psychological or psychophysiological research, nor is it convincing in its inner logic [13,22,35]. It is easily imaginable that an innocent suspect recognizes the relevant questions as most pertinent, and will therefore show large responses.[1] Simply put; the relevant and control questions differ on a number of dimensions besides deception, meaning any difference in psychophysiological responding cannot be solely attributed to deception.

The shortcomings of the CQT were recognized in the late fifties of the previous century, amongst others by psychologist David Lykken. Lykken developed an alternative question format, which he named the Guilty Knowledge Test (GKT; [33,34]). This test is nowadays commonly referred to as the Concealed Information Test (CIT; Verschuere, Ben-Shakhar, & Meijer, 2011). In contrast to the CQT, the CIT does not measure deception, but attempts to establish whether an examinee possesses pertinent crime related information. In the CIT, questions presented to the examinee (e.g., 'the murder weapon was a′) are followed by one relevant alternative (e.g., the actual murder weapon: an ice pick) and several neutral (control) alternatives (e.g., a knife, a letter opener, a pair of scissors, a piercer) presented in random order. These neutral alternatives are chosen such that an innocent suspect would not be able to discriminate them from the relevant alternative. In contrast, a suspect who is familiar with the details of the crime would be able to discriminate between the relevant and the neutral control items, and the relevant items will elicit enhanced physiological responses such as increased skin conductance, a decrease in respiration, and changes in heart rate [36]. In sum, knowledge is inferred from systematic stronger responding to the correct alternatives.

[1] In as case in which one of the authors acted as an expert witness, the suspect had failed a CQT polygraph test. In court, the suspect noted about the relevant questions: 'those are the questions you are there for. No wonder the machine goes of. I experienced innocent stress.'

The CIT countered some of the main criticisms of the CQT, most notable the risk of an innocent suspect failing the test (i.e., false positive outcome). Under the assumption that all alternatives are equally plausible, an innocent suspect cannot distinguish between the relevant and the neutral control alternatives, and the false positive rate is expected to follow the laws of probability [35]. The probability of an innocent suspect showing – by chance – the largest response to the correct alternative in one question with five options is expected to be .2. The probability of this happening in three questions is expected to be $.2^3$ =.008, so less than 1%, and one can set the false positive rate at an arbitrary low level by using a sufficiently large number of questions (e.g., 5) and an adequate criterion for inferring guilt (e.g., show the largest response to the correct alternative on at least 4 out of the 5 questions).

Whereas the CIT can be used to detect an examinee's knowledge of crime-related details, when it comes to studying the construct of deception, the test is confounded: knowledgeable participants respond truthfully on the majority of trials – namely presentation of the neutral control alternatives that typically constitute 80% or more of the trials -, while being deceptive only on the minority, i.e., only upon presentation of the relevant alternative [37]. For the purpose of detecting concealed information, this is not problematic, as only the knowledgeable participant can discriminate the relevant from the neutral controls. But, for the scientific study of deception, the CIT is problematic because besides to deception, any differences in responding can also be attributed to a frequency effect.

A third questioning format was developed to specifically study deception by isolating the deceptive response. This paradigm was originally developed by John Furedy and his colleagues using skin conductance (e.g., [14]), and named the Differentiation of Deception test (DoD). In the DoD, examinees are presented with a series of questions and are instructed to give truthful answers to half of them and deceptive answers to the other half. Alternatively, in a more recent variant of the DoD, participants are asked to answer each question twice: once truthfully and once deceptively. This test was labeled the Sheffield Lie Test (e.g., [48]). Because each question is answered both truthfully and deceptively, the DoD isolated deception to a high degree.

3. From polygraph to brain imaging

The CQT is fundamentally flawed because the relevant and control question differ on many other dimensions besides deception. The CIT is a valid test to detect knowledge, but not to study or detect deception. The DoD does isolate deception to a high degree. It should therefore not come as a surprise that the fMRI-based lie detection test we started this article with relied on a variant of the DoD paradigm. In fact, the rationale for this test (and its conclusion) was based on the findings of an earlier study by Spence and colleagues [48]. In this earlier study, Spence invited 10 participants, and presented them with a total of 36 autobiographical statements, such as 'made your bed' and 'taken a tablet' while the fMRI scanner registered their brain activity. Participants answered with key presses labeled 'yes' and 'no' based on color-coding (e.g., lie in response to green or red). Results revealed that lying – compared to truth telling - was associated with activity in the right ventrolateral prefrontal, left ventrolateral prefrontal and medial premotor area's. Ventrolateral prefrontal activation had previously been shown associated with response inhibition [18]. This pattern – which has been found in many studies since (for reviews see [15,16]) - led the authors to conclude that deception constitutes an executive function, including withholding the truth, and response manipulation and monitoring. In other words, the truthful responding is the default modus of the brain, and when being deceptive, this truth needs to be inhibited, and the deceptive response needs to be selected and executed.

The test of the 42-year-old woman revealed a pattern of brain activation highly similar to that found in the 2001 study: increased