The Comparative Advantages of Brain-Based Lie Detection: The P300 Concealed Information Test and Pre-Trial Bargaining

John Danaher*

Forthcoming in *International Journal of Evidence and Proof* (2015) 19 (1)

The lie detector test has long been treated with suspicion by the law. Recently, several authors have called this suspicion into question. They argue that the lie detector test may have considerable forensic benefits, particularly if we move past the classic, false-positive prone, autonomic nervous system-based (ANS-based) control question test, to the more reliable, brain-based, concealed information test. These authors typically rely on a "comparative advantage" argument to make their case. According to this argument, we should not be so suspicious of lie detection evidence if it has comparative advantages over the epistemic methods currently utilised by the legal system. In this article, I add to this growing support by making a novel comparative advantage argument in favour of brain-based lie detection evidence. The argument focuses on the P300 concealed information test (P300 CIT), which has several unique properties, and on the effect it may have on pre-trial bargaining in criminal cases rather than in-court evidence. The thesis is that P300 CIT could allow for innocent defendants to credibly signal their innocence to investigators and prosecutors during pre-trial bargaining more effectively than current proposed methods for doing the same thing. I defend this argument from a number of objections, and suggest that it opens up an interesting avenue in the ongoing debate about the merits of this technology. Although the argument is presented with the criminal law in mind, it could form part of a more general cumulative case in favour of this technology.

Keywords: Concealed Information Test; Lie Detection; Neuroscience; Plea Bargaining; Scientific Evidence

1. Introduction

There has been a lot of hype in recent years about brain-based lie detection.[1] If its purveyors are to be believed, new technologies and new experimental results are bringing us tantalisingly closer to a scientifically sophisticated kind of mind-reading. Such a thing would have many obvious forensic benefits. Courts rely heavily on witness testimony to determine the facts of a case, but witnesses can be deceptive and misleading. We can try to incentivise them to tell the truth by imposing penalties, but this assumes we can know whether or not they are telling the truth in the first place. A scientific test that allowed us to bypass all this uncertainty could be a real boon.

Despite this, legal systems, particularly in the US, have long been suspicious[2] of such tests. One can easily see why. If technologies of this sort are unreliable, or if they depend on invalid methods for inferring deception, they can be easily abused and increase the number of miscarriages of justice. Furthermore, tests of this sort seem to trespass upon a domain that is quintessentially within the competence of the court: determining the credibility of witnesses.

Is this suspicious attitude always warranted? Recently, some legal theorists have argued that it is not.[3] In doing so, they tend to highlight both the increased reliability of the brain-based tests, and, more importantly, the *comparative advantages* of such tests over the epistemic methods currently employed by the legal system. This article aims to add weight to this view by developing a novel comparative advantage argument in favour of brain-based lie detection. It does so by focusing on a particular kind of test — the P300 concealed information test

---

* School of Law, National University of Ireland, Galway

[1] The tangible evidence for this is probably the existence of companies, both U.S.-based, offering fMRI lie detection. For example No Lie MRI (http://www.noliemri.com) and Cephos Corp (http://cephosdna.com). The latter seems now to specialise in Forensic DNA evidence, though its founder, Dr. Steven Laken, has presented fMRI lie detection evidence in court. For example in the case *United States v Semrau, No. 07-10074 M1/P* (W.D. Tenn. May 31, 2010)

[2] Though the technology has been embraced elsewhere, such as in the military and in employment. See: K. Adler *The Lie Detectors: The History of an American Obsession* (New York: Free Press, 2007).

[3] J. Meixner 2012. 'Liar, Liar, Jury's the Trier' (2012) 106(3) *Northwestern Law Review* 1451-1488; F. Schauer, 'Can bad Science be Good Evidence? Neuroscience, Lie Detection and Beyond' (2010) 95 *Cornell Law Review* 1191 -1219; and F. Schauer 'Lie Detection, Neuroscience, and the Law of Evidence' (2012)*Virginia Public Law and Legal Theory Research Paper 2012-49*

(CIT) — and on the phenomenon of plea bargaining in criminal cases. The claim is that the strategic use of the P300 CIT at this stage in the criminal justice process could help resolve the so-called "innocence problem" that plagues the plea bargaining system. This is the problem whereby, following standard assumptions about bargaining in the shadow of the law and rational choice, even innocent defendants should, if they wish to avoid the possibility of a greater punishment following a trial, plead guilty to some offence. The argument advanced is that optional use of the P300 CIT could address this problem more effectively than current proposals. This argument highlights the exculpatory and sometimes neglected pro-defendant nature of this technology. Furthermore, although the focus in this article is on pre-trial bargaining in criminal cases, potential uses of this technology in alternative forms of pre-trial bargaining are discussed, and the emerging cumulative case for the use of this technology is emphasised.

The remainder of the article proceeds in five parts. First, it explains exactly what the P300 CIT is, how it differs from other methods of lie detection, and why it is a promising technology. Second, it reviews the current comparative advantage arguments in favour of brain-based lie detection. Third, it develops a novel, pre-trial, plea bargaining, comparative advantage argument in favour of the P300 CIT. Fourth, it considers a number of objections and replies to this argument. And fifth, and finally, it concludes by considering the broader implications of this argument for the use of this technology in the legal system.

2. What is the P300 Concealed Information Test?

This article cannot pretend to offer a comprehensive summary of the current science behind the P300 CIT,[4] but some essential background information must be presented. To that end, this section briefly covers the history of the CIT, the differences between it and the more-widely known control

[4] On the P300 CIT, see JP Rosenfeld, Hu Xiaoqing, E. Labkovsky, J. Meixner, and M. Winograd 2013. 'Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information' (2013) 90(2) *International Journal of Psychophysiology* 118-134 available at http://dx.doi.org/10.1016/j.ijpsycho.2013.08.012. On the CIT more generally, with some specific input on the P300 test, see B. Verscheure, G. Ben-Shakhar and E. Meijer *Memory Detection: Theory and Application of the Concealed Information Test* (Cambridge: Cambridge University Press, 2011)

question test (CQT), and the current research being done on the P300 version of the CIT. The goal is to convince the reader that the P300 CIT is a credible scientific test, which is less vulnerable to the problems that plague traditional forms of brain-based lie detection, and which, pending more extensive field testing, could readily be used in forensic settings.

Cinematic and other pop cultural depictions of lie detection often focus on the CQT.[5] The classic polygraph lie detection test adopts this format. A suspect is hooked-up to a machine that records his/her physiological response profile in the autonomic nervous system (ANS - the system that mediates the body's fight or flight response, among other things). The suspect is then asked a series of questions. These questions typically fall into three categories: (i) *control*, *i.e.* general questions that most people would be inclined to lie about; (ii) *relevant*, *i.e.* those specifically relevant to the incident under investigation; and (iii) *irrelevant*, *i.e.* those with no bearing on the matter under investigation. The principle underlying the CQT is that deceptive test subjects will exhibit a consistently larger physiological response profile to relevant questions than to control questions; innocent suspects will do the reverse. This makes it possible to tell the liars from the truth-tellers. Despite its popularity, the CQT has been repeatedly criticised for being improperly scientific, easily manipulated by the tester, and often adopted simply as an interrogation prop to force a confession.[6]

The CIT is a completely different type of test. The CIT does not directly test for deceptiveness (though it will tend do so indirectly).[7] Instead, it tests for the *recognition* of relevant information. Protocols for the CIT can vary, but in broad outline the subject will be presented with different kinds of stimuli, some of which are relevant to the matter under investigation (usually called the "probe" stimuli), some of which are not (but which can belong to a similar category of stimuli). A heightened physiological response to the probe stimuli suggests that the subject recognises the information. It is easy to see how this could form the basis of a

[5] Information about the CQT is drawn from: G. Ben-Shakar 'A Critical Review of the Control Questions Test' in M. Kleiner (Ed) *Handbook of Polygraph Testing* (London: Academic Press, 2002); J. Furedy and R. Heselgrave 'The Validity of the Lie Detector: A Psychophysiological Perspective' (1988) 15 *Criminal Justice and Behavior* 219; and National Academies of the Sciences Report, *The Polygraph and Lie Detection* (Washington, DC: National Academies Press, 2003).

[6] Sources in previous note on all these points.

[7] This is for two reasons. First, establishing recognition of information through a test would be unnecessary if suspects honestly admitted to recognising it. And second, because the test protocol itself sometimes indirectly assesses for deception. On this, see: D. Meegan, D. 'Neuroimaging Techniques for Memory Detection' (2008) 8 *American Journal of Bioethics* 9-20. There is some dispute within the scientific community as to whether indirectly testing for deceptiveness improves the accuracy of the test. Rosenfeld et al 2013 (n 4) review this debate, see section 4.1 of their article.

forensically useful test. For example, in the case of a criminal investigation — which is where the focus shall be for the remainder of this article — a suspect could be tested for their recognition of crime-relevant details, *e.g.* murder weapon, venue, victim's dress and so on.

For the CIT to work there must be some detectable physiological signal that reliably indicates whether or not the subject recognises the information being presented. ANS signals can be used for this, and indeed they have been and are being used in certain parts of the world.[8] The problem with these signals, however, is their vulnerability to *countermeasures*, *i.e.* techniques employed by test subjects to confound or undermine results of the test.[9] Brain-based signals are more promising, at least to the extent that they involve subconscious processing, as it is less likely that a test subject could consciously manipulate the patterns being detected.

This is where the P300 CIT comes into play. The P300 is a particular kind of brainwave ("evoked response potential") that can be detected with the use of an electro-encephalograph (EEG) monitor. The brainwave is evoked whenever a subject is presented with a rare and meaningful stimulus. Information related to a crime, that is known only to the test subject and the investigator, would be a paradigm example of such a stimulus.[10] Little wonder then that the P300 has been proposed for use in forensic CITs.

An early proponent of the P300-based CIT was Lawrence Farwell. He was involved in some of the initial mock crime tests of the technique.[11] These tests suggested that very high accuracy rates (>90%) could be achieved with the P300 measure, with errors tending in favour of false negatives rather than false

---

[8] It is routinely used as part of the investigative process in Japan. For a discussion see: A. Osuki 'Daily application of the concealed information test: Japan' in B. Verscheure, G. Ben-Shakhar, and E. Meijer (eds) *Memory Detection: Theory and Application of the Concealed Information Test* (Cambridge: Cambridge University Press, 2011)

[9] For example, see G. Ben-Shakar and K. Dolev 'Psychophysiological detection through the guilty knowledge technique: effects of mental countermeasures' (1996) 81(3) *Journal of Applied Psychology* 273-281; and E. Elaad and G. Ben-Shakhar 'Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test' (1991) 11(2) *International Journal of Psychophysiology* 99-108. References were originally sourced through Rosenfeld et al. 2013 (n 4), which contains a lengthy discussion of countermeasure susceptibility in the P300 test.

[10] This means that there cannot be leakage of the information to innocent suspects. For a discussion of this see: MT Bradley, CA Barefoot, and AM Arsenault 'Leakage of information to innocent suspects' in B. Verscheure, G. Ben-Shakhar, and E. Meijer (eds) *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press, 2011).

[11] LA Farwell and E. Donchin 'The truth will out: interrogative polygraphy ("lie detection") with event related potentials' (1991) 28(5) *Psychophysiology* 531-547.

positives. In other words, the test was more likely to falsely label someone innocent than guilty — advantageous given the oft-cited normative preferences of the criminal law for exculpation over inculpation.[12] Following these initial experiments, Farwell developed his own, patented version of the test (sometimes referred to as "brain fingerprinting")[13] and received some reasonably high-profile media exposure for his efforts to get it accepted by the courts. Those efforts were successful in one case, with a court[14] admitting the evidence under the Daubert test,[15] though the technique proved less successful in a later case.[16] In subsequent years, Farwell involved himself in classified military and security service tests of his technique, before recently re-emerging to the academic world to argue for the strengths of his version of the P300 CIT.[17]

To say that Farwell's test has been controversial would be an understatement. The test and the claims he makes on its behalf have been repeatedly criticised by some of the leading figures in the field of psychophysiology,[18] and the general candour of his work sets flying too many red flags (media promotion prior to scientific acceptance; secret government tests; lack of proper peer review; publication in lower quality journals)[19] to be taken seriously.

---

[12] "Better that ten guilty go free than one innocent be found guilty".

[13] For general details, see Farwell's company webpage, formerly located at http://www.brainwavescience.com, but now at http://www.governmentworks.com/bws (visited 23/1/14). The unique feature of Farwell's test is that in addition to recording the P300 response it also focuses on something Farwell calls the MERMER response. Having been rather inactive in the academic community for a number of years (apparently due to military testing of his technique), Farwell has recently returned to publishing, with a number of research and review articles touting the strengths and weaknesses of his techniques. L. Farwell, D. Richardson, and G. Richardson 'Brain Fingerprinting Field Studies Comparing P300-MERMER and P300 brainwave responses in the detection of concealed information' (2013) 7(4) *Cognitive Neurodynamics* 263-299; also L. Farwell 'Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials' (2012) 6 *Cognitive Neurodynamics* 115-154.

[14] *Harrington v. State of Iowa* 659 N.W.2d 509 (Iowa 2003)

[15] The test should probably not have been admitted given the paucity of ecologically valid experiments that existed at the time. For a critique see JP Rosenfeld 'Brain Fingerprinting: a Critical Analysis' (2005) 4 *Scientific Review of Mental Health Practice* 20-37.

[16] *Slaughter v. State of Oklahoma* 105 P.3d 832 (Oklahoma 2005)

[17] See Farwell, Richardson and Richardson 2013 (n 13); and Farwell 2012 (n 13)

[18] For example: Rosenfeld 2005 (n 15); Meegan 2008 (n 7); and E. Meijer, G. Ben-Shakhar, B. Verschuere, and E. Donchin 'A Comment on Farwell (2012): brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials' (2013) 7(3) *Cognitive Neurodynamics* 155-158. The latter paper is co-authored by Donchin, who worked with Farwell on the original 1991 experimental tests of the P300 CIT. The critique ends with the following, rather damning, statement "By selectively dismissing relevant data, presenting conference abstracts as published data, and most worrisome, deliberately duplicating participants and studies, he misrepresents the scientific status of brain fingerprinting. Thus, [Farwell] violates some of the cherished canons of science and if [he] is, as he claims to be, a 'brain fingerprinting scientist' he should feel obligated to retract the article."

[19] See previous note, particularly Meijer et al for illustrations of these problems with Farwell's work.

A much more promising version of the P300 CIT, with far more scientific credibility behind it,[20] comes from work done by the Rosenfeld Lab in Northwestern University.[21] The research emanating from this lab is notable for two reasons. The first reason is that it systematically tries to address and resolve the problem of countermeasures in the P300 CIT. Although I earlier suggested that brain-based tests were less susceptible to countermeasures than autonomic tests, this is not entirely true. The classic P300 test-protocol — in which a suspect is presented with three different kinds of stimuli (probe, irrelevants, and targets[22]) — can be confounded by getting the test subject to perform attention forcing exercises.[23] One of the key developments from the Rosenfeld lab is a more complex test protocol (appropriately named the Complex Trial Protocol) that, so far, seems to be resistant to such countermeasures.[24] That said, it is not completely immune (a concept that is distinct from "resistance")[25] to countermeasures and researchers are working on ways on minimsing their effects.[26]

The second reason that research from this lab is notable is for its attempt to increase the ecological validity of the P300 CIT. A longstanding criticism of experimental tests of the technique is that they provide little insight into how such a test may operate in the real world. The lab is a controlled environment. Even in mock crime scenarios, the test subjects can rehearse the details of the crime, leakage of information between testers and subjects can be minimised, and the tests can take place shortly after the mock crime was committed. The real world might be very different. Criminals may not rehearse and meticulously plan the details of their crimes. Investigations could take some time, and so a suspect

---

[20] I base this partly on the number of experimental tests published; the high quality journals in which they are published; and the willingness on the part of the experimenters to admit to flaws and weaknesses in the test.

[21] http://groups.psych.northwestern.edu/rosenfeld/home.html

[22] A target stimuli is not-related to the event under investigation, but belongs to a similar category of stimuli. For example, if you were investigating awareness of a murder weapon which was, in fact, a handgun (probe) you might also present a shotgun or knife to the test subject. These would count as target stimuli.

[23] JP Rosenfeld, M. Soskins, G. Bosh, and A. Ryan 'Simple, effective countermeasures to P300-based tests of detection of concealed information' (2004) 41(2) *Psychophysiology* 205-209.

[24] Initially presented in JP Rosenfeld et al. 'The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for the detection of concealed information' (2008) 45(6) *Psychophysiology* 906-919. The CTP has been the basis of over a dozen peer-reviewed publications now, with several more on the way. See Rosenfeld et al. 2013 (n 4) for a review.

[25] See Rosenfeld et al 2013 (n 4), section 3 for a discussion. The point is that the test still distinguishes the true suspects from the false ones, even if they using countermeasures. So it is resistant in that respect.

[26] See M. Winograd and JP Rosenfeld 'Countermeasure mechanisms in the complex trial protocol' (2012) 85(3) *International Journal of Psychophysiology* 305; and J. Meixner and JP Rosenfeld 'Countermeasure mechanisms in a P300-based concealed information test' (2010) 47(1) *Psychophysiology* 57-65.

might not be questioned until several weeks (possibly even years) after the crime took place. Furthermore, if the details of the crime are widely-known, or if they are leaked to the press, then the potential suspect could innocently acquire knowledge of them, thereby confounding the results of the test.

Daubert-type admissibility criteria for scientific evidence pay close attention to the issue of ecological validity, and any proponent of the P300 CIT would need to address the issue before it could be admitted for forensic use. The Rosenfeld lab have tried to do this by performing experiments in which recognition of "incidentally-acquired" information has been tested,[27] in which there is up to a one-month time-lag between information acquisition and testing,[28] and in which mock crime scenarios are generally more realistic than was previously the case. In addition to this, the researchers have experimented with a variety of techniques for increasing the accuracy of the test.[29]

In general, the results of these tests are impressive. Reported accuracy levels vary from 75% to 90%, with the errors biased in favour of false negatives rather than false positives, and with accuracy increased if certain enhancements to the test protocol are used.[30] The suggestion now seems to be that this version of the P300 CIT simply needs extensive field-testing before it will be ready for systematic use in forensic contexts.[31]

This should give the reader a rough picture of where we now are with the P300 CIT. Based on this picture, and in particular on the work emanating from the Rosenfeld lab, my working assumption for the remainder of the article shall be: (a) that the CTP version of the P300 CIT is something that could soon be ready for systematic forensic use; (b) that the accuracy of the test ranges from 70-

---

[27] M. Winograd and JP Rosenfeld 'Mock crime application of the complex trial protocol P300-based concealed information test' (2011) 48(2) *Psychophysiology* 155-161

[28] X Hu, and JP Rosenfeld 'Combining the P300-complex trial-based Concealed Information Test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection' (2012) 49(8) *Psychophysiology* 1090-1100.

[29] See, for example: X. Hu et al 2011. 'Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection' (2011) 49(1) *Psychophysiology* 85-95; JP Rosenfeld, X. Hu and K. Pederson 2012. 'Deception awareness improves P300-based deception detection in concealed information tests' (2012) 86 *International Journal of Psychophysiology* 114-121; and X. Hu, N. Pornpattananangkul, and JP Rosenfeld 2013. 'N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection' (2013) 50(5) *Psychophysiology* 454-464.

[30] Previous footnote. Also Rosenfeld et al 2013 (n 4).

[31] See Meixner 2012 (n 3) and Rosenfeld et al 2013 (n 4), section 6.

90% (a generally a conservative estimate of accuracy); and (c) that it is more likely to falsely exonerate than falsely inculpate.

3. Lie Detection and Comparative Advantage

If the P300 CIT is to be used in forensic contexts, one thing it will have to confront is the longstanding suspicion of lie detection tests that I mentioned earlier.[32] It is possible that further experimental tests will overcome this resistance, particularly if those tests are sensitive to the standards demanded by the courts in Daubert-style admissibility tests. This is one strategy being employed by the researchers at the Rosenfeld lab.[33] But the current resistance may also derive from biases within the prevailing legal and policy-making culture, and from problems with the normative standards of the legal system. Consequently, alternative strategies of argumentation may need to be employed before the P300 CIT is deemed fit for forensic use.

This certainly seems to be the belief of Frederick Schauer and John Meixner,[34] both of whom suggest that resistance to brain-based tests of this sort derives, at least in part, from unwarranted biases and unchecked assumptions operating within the legal system. These include: mistaken beliefs about what the technology does; ignorance of the weaknesses of existing legal methods for acquiring the same kind of information; irrational fealty to historical norms of evidence gathering; and so on. Both authors employ "comparative advantage"-arguments in order to overcome these sources of resistance. Such arguments draw attention to the flaws in the epistemic methods currently employed by the legal system and then highlight ways in which the new technology, even if it is far from perfect, represents an improvement.[35] Since my goal is to develop a similar style of argument, a quick review of the arguments adopted by Schauer and Meixner is in order.

Schauer's argument is concerned with fMRI-based lie detection, which tends to adopt a CQT format as opposed to a CIT format.[36] Schauer suggests

[32] See Schauer 2010 (n 3) for a lengthy discussion
[33] Rosenfeld et al 2013 (n 4), section 6
[34] Meixner 2012 (n 3)
[35] One can look upon this as a way of overcoming status quo bias. See N. Bostrom and T, Ord, 'The Reversal Test: Eliminating Status Quo Bias in Applied Ethics' (2006) 116 *Ethics* 656-679
[36] Though fMRI methods could be used as the basis for a CIT too. See Meegan 2008 (n 7) on this; also M. Gamer 'Detecting of deception and concealed information using neuroimaging techniques' in B.

that the resistance to this method of lie detection may derive from a misunderstanding of the significance of its overall level of accuracy (which he also estimates at between 70-90%). His feeling is that critics believe that a test with a reasonably high probability of error like this should not be accepted by courts. As it happens, Schauer may be wrong in thinking that this can explain the resistance.[37] But that does not really matter here. What matters is the argument he builds to address this misunderstanding. That argument has two prongs. The first — which we may call the *probative context* prong — suggests that in some probative contexts a test with a 30% probability of error could still be highly valuable, e.g. in civil cases where the standard of proof is lower, or in criminal cases where the defence only has to raise a reasonable doubt. The second prong — which is the *comparative advantage* prong — points out that a lie detection test with a 70% probability of success is still a good deal better than existing legal methods for ensuring that a witness tells the truth. For example, the administration of the oath, perjury laws and robust cross-examination, all of which are designed to encourage truth-telling, have completely unknown error rates: a movement away from those methods to something that at least has a tractable probability of error would be an improvement.

Meixner's argument is concerned with the P300 CIT.[38] Indeed, in addition to being a lawyer, Meixner is a former researcher in the Rosenfeld lab. One of his suggestions is that resistance to the P300 CIT and other methods of brain-based lie detection, may be driven by the belief that the results of such tests trespass upon a matter that is within the exclusive prerogative of the court (judge or jury), namely: assessing the credibility of witness testimony. This suggestion has some basis in reality as courts have already rejected such tests on these grounds.[39] Meixner addresses this resistance with on two grounds. The first is that the P300 CIT is not, primarily, about determining witness credibility: it is about determining whether someone recognises a stimulus or not. This may, of course, have the effect of impugning a witness' credibility, but that's an indirect effect. The second argument is that, even if the P300 CIT did trespass upon this issue of witness credibility, there is no strong reason to think that it should not be admitted. This is because — as Meixner illustrates by reviewing a large swathe of

---

Verscheure, G. Ben-Shakar, and E. Meijer (eds) *Memory Detection: Theory and Application of the Concealed Information Test* (Cambridge: Cambridge University Press, 2011).

[37] The resistance may not be due to the high probability of error, per se, but rather a lack of certainty about the error rates in real world cases. Still, as Schauer 2012 (n 3) points out courts are willing to accept other methods with completely unknown error rates.

[38] Meixner 2012 (n 4)

[39] See the decision in *Wilson v. Corestaff Services LP* (2010) 900 NYS 2d 639

the experimental literature[40] — there is no reason to think that judges or juries are particularly good at assessing credibility. Indeed, the studies he reviews suggest that they hover around chance, and that they may issue a high number of false positives, particularly if they are trained in methods of behavioural lie detection (due to a false confidence effect). So, once again, the brain-based method offers us a comparative advantage over the existing system.

Schauer and Meixner's arguments could be criticised on a number of grounds.[41] I will not do so here as my goal is not to defend them, but to offer an alternative type of comparative advantage argument. Where both Schauer and Meixner focus on the benefits to the legal system at the trial stage, my argument focuses on the pre-trial stage. And where Schauer and Meixner are concerned with both the inculpatory and exculpatory powers of these tests, my argument is concerned primarily with the exculpatory powers. Consequently, I will suggest that my argument has merits that their two arguments lack. This, however, is not to suggest that my goals are irreconcilable with theirs. Far from it. Nothing I say here directly contradicts or undermines what they have to say. Indeed, if my argument is persuasive it could be joined with theirs to form a cumulative case in favour of the use of this technology. This may be an advantage. After all, resistance to this technology may be attributable to a number of independent factors. Consequently, it makes sense to counter that resistance along several fronts.

4. How might the P300 CIT improve pre-trial bargaining?

The argument I wish to defend is that granting suspects the optional use of the P300 CIT helps to address a signalling problem (the "innocence problem") that is inherent in systems of pre-trial plea bargaining. The argument is an adaptation of one originally defended by Russell Covey.[42] The difference

[40] Meixner 2012 (n 3), pp. 1461-1473

[41] For example, juries may be unreasonably persuaded by brain-based tests due to the imaging techniques associated with them. Schauer (2012 n 3) discusses this problem, and it may well be overstated — e.g. L. Klaming, 'Does Neuroscientific Evidence Bias Legal Decision-Making? Some Preliminary Findings' (2011) 3 *Law, Innovation and Technology* 303-317 — but it is probably worth exploring. Relatedly, Meixner's critique of juries ignores possible Condorcet effects that could improve the accuracy of jury assessments. Condorcet effects arise from jury theorems. These probability theorems suggest that groups of individuals voting on the truth of a proposition can be quite good at identifying the right answer, even if they are individually only marginally better than chance. I am indebted to David Duffy for drawing this point to my attention.

[42] R. Covey, 'Signaling and Plea Bargaining's Innocence Problem' (2009) 66 *Washington and Lee Law Review* 73-130.

between my argument and that of Covey is that where Covey suggests optional interrogation as an (existing) solution to the signalling problem, I suggest optional use of the P300 CIT. I submit that my proposal has significant comparative advantages over his. Furthermore, while the focus of my argument will be on bargaining prior to a criminal trial, it does not rule out the possible uses of the P300 CIT in other forms of pre-trial bargaining since they too may be vulnerable to something akin to the innocence problem.

The argument consists of five propositions. The first is that there is an "innocence problem" inherent in systems of plea bargaining. The second is that the innocence problem arises from a signalling problem. The third is that the signalling problem can be addressed by introducing "subwagers" into pre-trial bargaining. The fourth is that the optional use of the P300 CIT would constitute such a subwager. The fifth is that this solution is better than Covey's proposed solution.

The innocence problem is illustrated by the dominant theories of pre-trial bargaining. Following the classic presentation by Mnookin and Kornhauser,[43] it is common to say that pre-trial bargaining — across all forms of litigation — takes place in the "shadow of the law". That is to say: the possibility of legal sanction or punishment provides an ominous backdrop against which litigants can negotiate mutually agreeable pre-trial settlements. In the case of criminal prosecutions, these settlements take the form of plea-bargains.[44] Plea-bargaining is the common practice whereby someone who is charged with a criminal offence will agree to plead guilty to a lesser offence in order to avoid the risk of being tried for the more serious offence.

Adopting standard rational choice assumptions, plea bargaining should be attractive to both prosecutors and defendants. Prosecutors wish to maximise the amount of punishment per unit of prosecutorial resources; trials are costly (the resources spent in one trial could be used to secure other convictions); and their results are unpredictable. Accepting a lesser plea helps to maximise convictions and minimise costs. Contrariwise, from the defendant's perspective, the goal is to

---

[43] R. Mnookin and L. Kornhauser, 'Bargaining in the Shadow of the Law: the Case of Divorce' (1979) 88 *Yale Law Review* 950. See also R. Cooter, S. Marks, and R. Mnookin 1982. 'Bargaining in the Shadow of the Law: A Testable Model of Strategic Behavior' (1982) 11 *Journal of Legal Studies* 225

[44] An early analysis of this can be found in W. Landes, 'An Economic Analysis of the Courts' (1971) 14 *Journal of Law and Economics* 61. See also F. Easterbrook, 'Criminal Procedure as a Market System' (1983) 12 *Journal of Legal Studies* 289.

minimise the amount of punishment. But since trials are unpredictable, pleading to a lesser offence is often the best way to minimise expected punishment. Indeed, so strong are the incentives on the defendant's side, that submitting a guilty plea can even be the dominant strategy for a large number of innocent defendants. The argument for this is that if there is not overwhelming objective evidence of innocence, and if the prosecutor has discretion to offer any sort of deal, then there is some minimal degree of punishment that the defendant should be willing to accept in order to avoid the risk of trial.[45] This is plea bargaining's *innocence problem*. Covey suggests that although real-world defendants do not always match the predictions of rational choice theory, the problem seems to be reflected in the real-world data, which suggests that in the US approximately 95% of all felony criminal cases result in guilty pleas.[46]

The innocence problem itself arises from a signalling problem. To understand this we need to understand what a signalling system is and how problems can arise in such systems.

A "signalling system" can be defined as any system with these three elements:[47]

Message Set: A set of possible signals that could be sent through the system, e.g. "P" and "not-P". Typically, these signals are propositions about an actual or likely state of the world.

Sender: An agent who selects one of the possible signals from the message set and conveys it to another agent (the receiver).

Receiver: An agent who must interpret the signals from the sender and assign a truth value to them (e.g. "P is true").

Both the pre-trial bargaining system and the trial system are signalling systems. In each case, there is a set of signals (claims about or related to guilt or innocence), that are sent by one set of agents (witnesses, defendants etc.) to another set of

[45] Of course, this assumes that the prosecutor has discretion to offer minimal amounts of punishment. Mandatory sentencing requirements can upset this assumption. But this is sometimes used to critique mandatory requirements. The argument being that they prevent efficient dispositions of criminal cases.
[46] Covey 2009 (n 42), p. 74 citing D. Berman and S. Bibas 'Making Sentencing Sensible' (2006) 4 *Ohio State Journal of Criminal Law* 37, at p. 42.
[47] For a longer discussion, see R. Koppl 'Epistemic Systems' (2006) 2 *Episteme: a journal of social epistemology* 91-106.

agents (prosecutors/juries/judges), such that the latter must assign a truth value to the signals sent by the former.

There are many problems that can afflict a signalling system: the message set may be insufficiently detailed to convey the truth; the sender may not be able to "see" all the elements of the set; the quality of the signals may degrade as they are passed through the system; and so on. Despite these possibilities, for present purposes, the phrase "signalling problem" can be taken to refer to a scenario in which the receiver is incapable of distinguishing true signals from false signals.

This is what happens in the case of pre-trial plea bargaining. When there is not overwhelming objective evidence of innocence, a defendant might have private knowledge of his or her innocence, but no way to credibly signal this to the prosecutors. Declaring innocence will not do since guilty defendants are just as likely to proclaim their innocence as genuinely innocent defendants. Their claims contaminate the pool of signals being sent to prosecutors. The result is that signals are indistinguishable from the prosecutorial perspective. Hence, we have a signalling problem: the receivers cannot separate the true signals from the false signals.

To resolve a signalling problem like this, the structure of the pre-trial bargaining system must be changed so that the true signals can be distinguished from the false. That much is obvious. Covey suggests that this can be done by introducing a "subwager" into the bargaining game that is being played between prosecutor/investigator and defendant. This is an optional move within the game that will be attractive to genuinely innocent defendants, but not to those merely masquerading as innocent. He explains this concept using an analogy with a simple card game.[48] I shall outline this analogy in some detail since it helps to understand both how my argument is similar to that of Covey and how it differs in certain crucial respects.

The card game works like this. There are two players: (a) the Gambler, G; and (b) the House, H. The House deals four playing cards, face-down, onto the table in front of the Gambler. Two of the cards must be red and two must be black. The cards are dealt at random so neither H nor G knows what order they are in. If the bottom card is red, then G gets to keep his money. If the bottom is

[48] See Covey 2009 (n 42), pp. 103-110

black, G has to pay H £20. This means that G is either going to lose money, or break even; and H is either going to break even or gain £20. Since the probability of the bottom card being red is 0.5, the expected payoffs for G and H are minus £10 and plus £10, respectively. We assume that dealing out the cards is costly to H, and so she would rather stop the game before dealing out. Consequently, H is willing to offer G a deal: she will accept half of her expected payoff now (i.e. £5), and forgo playing the remainder game. From G's perspective this is a good deal: the expected loss from the deal is less than the expected loss from playing the game.

Obviously, the argument is that this is analogous to what happens during plea-bargaining. The gambler is like the defendant who risks losing a lot by going to trial; the house is like the prosecutor who would prefer to avoid the costs of going to trial; and the proposed offer is like a plea bargain. If the defendant (gambler) is rational he will typically prefer to accept the plea bargain rather than risk playing the game to completion.

We can add a twist to the game whereby G has something akin to private knowledge of innocence but no way to credibly signal it to H. This can happen if we allow G to turn over the top card, but prevent him from showing it to H. Once G knows what the top card is, the probability of the bottom card being red will change, and so too will his expected losses from the game. For example, if the top card is black, the probability of the bottom card being red increases from 0.5 to (approximately) 0.67. In an ideal world, G should be able to use this information to negotiate a more favourable pre-game deal with H. But G cannot do this as the rules of the game prevent him from credibly signalling what he has seen to H. Simply proclaiming that the top card was black will not be enough since if the top card was red, G would have an incentive to lie about it, and H would know this. This is like the innocent defendant who earnestly proclaims his innocence but finds his claims contaminated by a pool of similar proclamations from guilty defendants.

Covey argues that the addition of a subwager to the game can help to address the signalling problem. The subwager is like an additional gamble within the game, and changes what was previously a one-step game into a two-step game. As follows:

Step One - Private Information: H deals four cards face-down onto the table, two red, two black. G gets to look at the top card, but cannot show it to H. G holds the top card to one side.[49]

Step Two - The Subwager: H makes G a new offer, which he can either accept or reject. If he rejects the game continues as before. If he accepts, then a new game is played. The remaining three cards are shuffled and the identity of one of them is revealed to both parties. If that new card turns out to be red, H must settle the game for much less than what she was originally demanding (say £1.25). But if that card turns out to be black, then H can settle the game for much more than she was originally demanding (say £10).

Should G accept the subwager? Well, it all depends on what he saw when he turned up the top card during step one. If the top card was black, then he knows that there is 0.67 chance that the card revealed during the subwager will be red. Those odds look pretty good. Conversely, if the top card was red, then he knows that there is a 0.67 chance that the card revealed during the subwager will be black. That's not so good. The upshot of this is that if the top card was black, G has a strong incentive to accept the subwager. Conversely, if the top card was red, he has a strong incentive not to take the subwager.

Something interesting has now happened. By introducing the subwager, H has given G a way to credibly signal his private information about the first card in the pile back to H. G's willingness to accept the subwager is, in effect, a signal to H that the top card was black and vice versa. H could then use this information to negotiate an alternative settlement. The signals that were previously mixed are now more distinguishable.

It is important to think about how the subwager works as a signalling device since we are going to be arguing for the incorporation of an analogous subwager mechanism into the plea-bargaining system. The subwager works primarily because of G and H's knowledge of the probabilities within the game. Indeed, the signalling power of the subwager varies quite dramatically depending on those probabilities. For example, if in some variation of the game the probability of the bottom card being red after the subwager was 0.87, G's willingness to take the

[49] Technically, a third party may be needed to observe the cards at this stage to remember the original order since this gets shuffled around during the subwager. See Covey (n 42) p. 104, fn 126 on this point.

subwager would be an even more powerful signal of what he saw during the first stage of the game. If the probability were lower (e.g. closer to the original 0.5) it would be relatively useless. This is because, in order to function as an effective signal, the subwager must represent a serious risk to the player who saw a red card at the first stage of the game (the "guilty" defendant in our analogy), and a small risk to the player who saw a black card (the "innocent" defendant). If this asymmetry of risks does not obtain, the subwager will be deprived of its signalling power. This is important: one of the chief arguments below is that voluntary submission to P300 CIT is an ideal subwager because such an asymmetry of risks does arise in the administration of that test.

Another important point here is to highlight exactly what it is that is doing the "signalling" in the subwager game. The answer is that it is the decision to submit to the wager, and not the outcome of the wager itself: G's willingness to submit to the subwager signals what he saw during stage one; the wager itself has no signalling power since once it is completed the settlements for the game completely change. This is somewhat problematic since it suggests a way for Gs who saw a red card during the first stage of the game to disguise themselves as ones who saw a black card. How so? Well, they can simply forcefully insist upon their willingness to submit to the subwager and hope that this forceful insistence persuades H to settle the game for less without going through the full subwager process. This would be like bluffing in poker. Of course, this problem could easily be resolved by having H randomly complete certain subwagers, and thereby dissuade Gs who had seen a red card with the threat of a much worse settlement offer. Still, this weakness in the signalling power of the subwager does suggest that alternative subwagers, ones that do not rely entirely on the willingness to gamble as their signalling mechanism, would be preferable. To be more precise, subwagers in which *both* the willingness to gamble, *and* the outcome of the subwager had some signalling power, would be better than the mechanism used in this card game. Again, this is significant insofar as one of the key arguments below is that the optional use of P300 CIT has this dual-signalling power.

The question is whether anything like the subwager (with equivalent or greater signalling power) could be introduced into the pre-trial bargaining game in criminal cases. Covey argues that something could be: the current system in the US whereby suspects can voluntarily submit to interrogation can function as

a subwager, albeit only if there is robust protection of the right to silence.[50] The argument works like this: If the protection of the right to silence is robust — i.e. if it really is true that guilt will not be inferred from silence — then a guilty defendant has a strong incentive not to submit themselves to questioning, for if they are guilty they are, presumably, more likely to be found out by submitting to interrogation: consistently maintaining a lie under adversarial conditions is more difficult than consistently maintaining the truth. Contrariwise, if they are innocent, they will have a good chance of establishing this through the interrogation: it's easier to maintain a consistent story and persuade the interrogators. Thus, their willingness to undergo interrogation, *coupled with the results of that interrogation*, provide them with a means of credibly signalling their innocence. The analogy with the subwager card game is almost direct, only now we have a subwager with dual signalling power. The optional use of interrogation creates a two-step game. The first step covers the arrest and initial threat of prosecution. At this stage, the defendant may have private knowledge of innocence but cannot credibly signal this to the prosecutor. The second step introduces the option of submitting to interrogation, which can be used to credibly signal innocence to the prosecutor.

Covey's argument is certainly ingenious, and he introduces some empirical evidence to suggest that innocent defendants are more likely to submit to voluntary interrogation,[51] and that this does correlate with a reduced likelihood of being charged.[52] Still, there are several problems with it. For one thing, the respective probabilities of innocent and guilty defendants making it through interrogation are unknown. In addition to this, the structure of the incentives, and the power of interrogation to get at the truth, need to be delicately balanced if it is going to work. It is far too easy to upset that balance and denude optional interrogation of its signalling power. For example, if the interrogation methods are too robust, then the likelihood of an innocent defendant holding up under the pressure will presumably diminish and it will be too easy to extract false confessions. This will make interrogation too costly from the defendant's perspective, and erode the signalling advantage. Similarly, it is quite possible that

[50] Why? Covey adopts the argument from Seidmann and Stein holding that the right to silence helps the innocent. For an overview, see A. Stein, 'The Right to Silence Helps the Innocent: A Response to Critics' (2008) 30(3) *Cardozo Law Review* 1115 - 1140

[51] S. Kassin and R. Norwick 2004. 'Why People Waive Miranda Rights' (2004) 28(2) *Law and Behavioral Science* 211-221.

[52] R. Leo, 'Inside the Interrogation Room' (1996) 86(2) *Journal of Criminal Law and Criminology* 266-303; and P. Cassell and B. Hyman 'Police Interrogation in the 1990s: An Empirical Study of the Effects of Miranda' (1996) 43 *UCLA Law Review* 839

hardened criminals will be able to hold up under robust questioning. If so, their presence within the pool of willing interrogatees will spoil the value of the signal.

I suggest that instead of voluntarily submitting to interrogation, the defendants be given the option of voluntarily submitting to a P300 CIT. The argument on behalf of this is straightforward: if you are guilty, and you really do have knowledge of key elements relating to the crime, you will have a strong incentive not to volunteer for the P300 CIT; conversely, if you are innocent, and know no crime-relevant concealed information, you will have a strong incentive to volunteer for the P300 CIT. Thus, the optional use of the test, coupled with its results, gives the innocent suspect the means to distinguish themselves from the guilty suspect. It does so with the dual signalling power mentioned above, and in a manner that is less susceptible to the problems associated with interrogation. The P300 CIT is a scientific test, with tractable probabilities of success, that has to be administered in accordance with strict protocols. It is not open to the kind of abusive strong-arming tactics associated with interrogation. As a result, the costs of submitting to the test will be more consistent and not high enough to disincentivise an innocent defendant from taking it. At the same time, the CTP version of the test, with its resistance to countermeasures, makes it less likely that hardened criminals will be able to contaminate the value of the signal. The result is that the introduction of the P300 CIT subwager would represent an improvement to the current system of pre-trial bargaining.

This gives us a comparative advantage argument in favour of the use of the P300 CIT. To briefly summarise, the argument consists of the following propositions:

> There is an innocence problem inherent in existing systems of plea-bargaining.
>
> The innocence problem arises from a signalling problem: signals sent by innocent defendants are indistinguishable from the signals sent by guilty defendants.
>
> Introducing a subwager into the pre-trial bargaining game can resolve this signalling problem by giving those with private knowledge of innocence a credible way to distinguish themselves from others.

Giving defendants the option of submitting to a P300 CIT provides them with just such a subwager.

The P300 CIT subwager is better than alternative subwagers, such as Covey's optional interrogation system.

Now that we have the argument in outline form we can proceed to address a number of objections and replies.

5. Objections and Replies

There are several different objections one could lodge against this argument. I shall group these into two families.[53] The first group argues that the optional use of the P300 CIT does not really represent a *useful* subwager in the pre-trial bargaining game; and the second group argues that it does not represent the *best available solution* to the innocence problem.

5.1 *Is the P300 CIT a useful subwager?*

There are two objections to contend with here. The first, and more serious, is something I call the *false negative problem*. The second, and less serious, is concerned with the possibly limited number of cases in which the P300 CIT could be used.

The false negative problem holds that the P300 CIT cannot be a genuine subwager because the test is biased in the direction false negatives. As mentioned earlier, in order for something to represent a genuine subwager, there must be some risk attached to it such that guilty defendants are going to be dissuaded from accepting it. In other words, taking the subwager must represent a genuine risk of inculpation to the guilty defendant. If this risk is minimal, then guilty defendants will be willing to take the subwager and thereby contaminate the signals being sent to the prosecutors. This might be the case if the P300 CIT had a high number of false negatives. For then, a guilty defendant could risk

[53] One could also argue that the innocence problem is not genuinely worthy of our attention. I ignore this possible criticism.

submitting themselves to the test, knowing that there is a good chance they would be not be implicated by its results.

The problem is that this is arguably true of the P300 CIT. Results from the Rosenfeld lab suggest that the number of false negatives can be as high as 30%, and results from other labs have obtained even higher rates of false negatives. For example, Allen and Mertens performed an experiment involving virtual reality mock crime scenario with a false negative rate of more than 50%.[54]

There are two responses to this problem. The first is that the extremely high false negative rate in the Allen and Mertens study may be anamolous. Rosenfeld et al argue that the study relied on a protocol in which rehearsed information was tested alongside incidentally acquired information, which reduced the sensitivity of the test.[55] Additionally, they submit that the use of the virtual reality mock crime, as opposed to a real world mock crime may have lowered the overall accuracy. In other mock crime tests, the false negative rate is lower and could be rendered even lower again by adopting the enhancements recommended by Rosenfeld and his colleagues.

More important than all this, however, is the fact that a high false negative rate is typically associated with a much lower false positive rate. Indeed, in the Allen and Mertens study there were no false positives at all. The coupling of a high false negative rate with a low false positive rate has some interesting properties when it comes to the subwager argument. It means that although there is a decent chance of a negative or inconclusive result, which will be exculpatory in nature; there is also a decent chance of a positive result which will be highly inculpatory in nature, given the extremely low rate of false positives. Thus, a guilty defendant who is deciding whether or not to take the test must balance the possibility of an exculpatory false negative against the risk of a highly inculpatory true positive. These risks are clearly asymmetric: the guilty defendant has much more to lose by submitting themselves to the test and possibly garnering a true positive, than they do by avoiding the test. It is submitted that this asymmetry of risks means that guilty defendants will not submit to the test and hence the signalling power of the P300 CIT subwager will be undiminished. This is not to mention the fact that the inculpatory/exculpatory balance of the test will

[54] R. Mertens and JJB Allen, 'The role of psychophysiology in forensic assessments: Deception detection, ERPs and virtual reality mock crime scenarios' (2008) 45 *Psychophysiology* 286–298
[55] See Rosenfeld et al 2013 (n 4), section 2

continue to make it attractive to innocent defendants (due to the low risk of false positives) and prosecutors (due to the high probative value of a positive result).[56]

The second member of this family of objections holds that the P300 CIT subwager will be ineffective because there are relatively few cases in which the test could be used. After all, the test can only be used in cases in which there is information about the crime that is known to the investigators and the guilty party alone, but not to anyone else (*i.e.* there is no information leakage); and in which there is a relatively short time lag between the administration of the test and the original crime.

There are certainly limitations to be acknowledged. The optional use of the P300 CIT will indeed be restricted to a certain class of cases, and strict protocols will need to be in place for its use. The test would need to be used relatively early on in the process. This is because existing evidence related to the effectiveness of the P300 CIT can only vouch for accuracy up to one month after an incident has taken place.[57] Furthermore, since the test checks the suspects *recognition* of crime-relevant details, the testers need to ensure that the suspect could not have learned of those details by innocent means (e.g. because the police already disclosed them to the suspect during questioning, or because the details were released to the media). Without this precaution, the results could be contaminated. This means that the details in question must be presented to the suspect for the first time during the test.

Despite the reality of these limitations, the objection as whole strikes me as being a less serious objection for at least two reasons. The first is that even if the utility of the P300 CIT is limited to such cases that is nothing to be sniffed at. If it represents a comparative advantage over pre-existing methods in only those cases where this protocol can be followed, it still represents a comparative advantage that we should embrace. The second reason is that it is not clear that this is a

[56] As with Covey's point about the need for a robust right to silence, it is important that the decision not to take the test is not admissible as evidence against the defendant. Otherwise, a refusal to take the test would be taken to automatically signal guilt, and guilty suspects would have an incentive to submit to the test (despite its risks).

[57] As mentioned in section 2, there are some studies covering a one month time-lag. Beyond this the reliability of the test is unknown, but some experiments suggest that the test cannot distinguish true memories from false memories, so caution is certainly needed: JJB Allen and R. Mertens, 'Limitations to the Detection of Deception: True and False recollections are Poorly Distinguished using an Event-related Potential Procedure' (2009) 4(6) *Social Neuroscience* 473-490 Given that the probability of false memories goes up over time, this provides some reason for limiting its use to the one-month time window (for the time being at least).

significant limitation. There will often be information about the crime that is only shared between guilty party and investigator, and the routine use of a P300 CIT in criminal investigations and prosecutions could alter practices so that the option of using the test is always introduced early on in the process, and so that possibility of information leakage is minimised.[58]

5.2 *Is the optional use of the P300 CIT the best solution to the innocence problem?*

Even if the P300 CIT subwager is a better solution to the innocence problem than traditional methods of interrogation, that does not mean that it is the best solution to the problem. Of course, it is very difficult to establish something as being the *best* solution to a problem. The most one can hope for is establishing something as being the best *for the time being* or the best *given the current pool of alternatives*. But even still one might worry that I have not considered all the available alternatives in my initial presentation of the argument.

In my rush to recommend the P300 CIT as a potential subwager, I passed over another set of tests that could perform much the same function.[59] I speak here of other forms of lie detection. For example, fMRI lie detection tests or even more traditional autonomic nervous system-based lie detection tests. Voluntary submission to tests of these sort could, arguably, provide a good way for innocent defendants to credibly signal their innocence.

In principle, this is true. In practice, there are some reasons for thinking the P300 CIT is still a better solution to the problem. First, to the extent that the alternative tests rely on a CQT format, it is not clear that they represent any deviation from the interrogation-solution proposed by Covey. As has repeatedly

---

[58] Japanese police forces have routinely used the older, ANS-based CIT, though not with the bargaining rationale that I suggest. This suggests that routine integration of its use into the criminal justice system would be possible. See Osugi, A. "Daily Application of the Concealed Information Test: Japan" in Verschuere, B., Ben-Shakar, G. and Meijer, E. *Memory Detection: Theory and Application of the Concealed Information Test* (Cambridge: CUP, 2011)

[59] In addition to this, in his original discussion, Covey noted that there are several other candidate solutions to the innocence problem. These include: the complete elimination of plea bargaining, the dismissal of all charges that fall below a certain evidential threshold, and improved pre-trial discovery mechanisms. I do not wish to dwell on these alternative solutions here as I think Covey does a reasonable job of refuting them (in brief, he argues that these solutions can seriously reduce the likelihood of convicting the guilty, and/or make things worse off for innocent defendants. See Covey 2009 (n 42), section II.B

been pointed out by its critics,[60] the CQT still relies on a robust, interrogatory method of questioning, in which the emotional salience of the questions is manipulated by the tester.[61] Indeed, in many cases the test functions as little more than an interrogation prop: a seemingly sophisticated device for inducing the subject to confess. Couple this with the relatively high level of false positives associated with the CQT and you have an option that will seem less attractive to the innocent defendant and less able to function as an effective subwager. In addition to this, there is the problem that ANS-based versions of either test are more vulnerable to countermeasures. The hardened criminal could leverage this fact to their advantage and contaminate the value of voluntarily submitting to this test.

6. Conclusion

In summary, although further field-testing certainly needs to take place, the legal system should be open to the use of the P300 CIT. One reason for this is that a test of this sort could provide innocent defendants with a means of credibly signalling their innocence to prosecutors and investigators during pre-trial bargaining in criminal cases. Furthermore, it can do this in a more effective manner than existing methods of signalling during pre-trial bargaining. And although the focus of this article has been on the use of the P300 CIT in that specific context, we can conclude by considering the broader implications of the argument.

We can do this in two ways. First, by asking whether the optional use of the P300 CIT could perform a similar signalling function in bargaining prior to non-criminal trials. The answer is yes, but with two important caveats. The P300 CIT only works when the recognition of information is of some significant probative value in the legal dispute between the parties. This is often the case in criminal trials. Whether it is often the case in civil (e.g. tort) trials is another matter. Certainly, one can imagine cases where it could be – for example, the recognition of an internal research memo on the dangers of certain industrial outputs in a toxic tort case – but recognition of this sort may not always be highly probative in those contexts. Furthermore, one has to bear in mind that the P300 CIT is only effective if the information that is being recognised has not been innocently

[60] Furedy and Heselgrave 1988 (n 5) and Ben-Shakhar 2002 (n 5).
[61] JJ Furedy 1993. 'The 'control' question 'test' (CQT) polygrapher's dilemma: logico-ethical considerations for psychophysiological practitioners and researchers' (1993) 15(3) *International Journal of Psychophysiology* 263-267.

acquired (the information leakage problem). It could be relatively easier for criminal investigators to control the leakage of crime-relevant information than it is for civil litigants to control the leakage of, say, tort-relevant information. This again could limit the utility of the optional use of the P300 CIT outside of criminal context. So although I would not exclude the application of my proposal to other contexts, the criminal trial has certain properties that make it uniquely favourable to it.

The second question is whether the utility of the P300 CIT is limited to the pre-trial bargaining phase. Certainly that is not the implication of the argument being advanced here. If the technology has the merits suggested earlier in this article, then it could be used more widely. My goal has simply been to highlight one area where the technology has particular advantages over existing epistemic methods employed within the legal system. As I said, this could be paired with arguments from the likes of Schauer and Meixner, and used as part of a cumulative case in favour of the more widespread use of the technology. That, however, is a project for another time.