Received: 15 April 2022 | Revised: 26 January 2023 | Accepted: 31 January 2023

DOI: 10.1002/acp.4050

RESEARCH ARTICLE

WILEY

Classification accuracy of the event-related potentials-based *Brain Fingerprinting* and its robustness to direct-suppression and thought-substitution countermeasures

M. Usman Afzali[1,2] | Richard D. Jones[1,3,4,5] | Alex P. Seren-Grace[1] | Robin W. Palmer[6] | Dena Makarious[1] | Mariana N. B. Rodrigues[1] | Ewald Neumann[1,7]

[1]School of Psychology, Speech and Hearing, University of Canterbury, Christchurch, New Zealand

[2]School of Psychology, Victoria University of Wellington, Wellington, New Zealand

[3]School of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand

[4]School of Medicine, University of Otago, Christchurch, New Zealand

[5]Christchurch Neurotechnology Research Programme, New Zealand Brain Research Institute, Christchurch, New Zealand

[6]School of Law, University of Canterbury, Christchurch, New Zealand

[7]New Zealand Institute of Language, Brain, and Behaviour, University of Canterbury, Christchurch, New Zealand

Correspondence
M. Usman Afzali, School of Psychology, Speech, and Hearing, University of Canterbury, Private Bag 4800, Christchurch, 8041, New Zealand.
Email: usman.afzali@canterbury.ac.nz

Abstract

Research on the accuracy of Brain Fingerprinting (BFP) has produced mixed outcomes: some report 99.9% and others report lower. Furthermore, no studies have measured the susceptibility of BFP to countermeasures. In Experiment-1, we report the accurate classification of 15 of the 16 subjects, tested on their own real-life autobiographical incidents; and 14 of the 15 other subjects, tested on another subject's real-life autobiographical incidents. In Experiment-2, 16 subjects of Experiment-1, who were tested on their own real-life incidents, participated in the BFP test again, but this time employing either direct-suppression or thought-substitution ($n = 8$ each) countermeasures. We report that neither direct-suppression nor thought-substitution was effective at concealing information that BFP was designed to reveal. We assert that BFP is a highly accurate, albeit not perfect, concealed-knowledge detection technology and that it is resistant to memory suppression and thought substitution countermeasures in the context of autobiographical incidents.

KEYWORDS
autobiography, Brain Fingerprinting, countermeasures, event-related potential, investigation

1 | INTRODUCTION

Recent advances in technology have resulted in the invention and development of forensic tools to fight crime. Some of these tools rely on the body's physiological responses (American Psychological Association, 2004), while others are based on reaction time (Verschuere et al., 2014), electroencephalograms (e.g., Brain Electrical Oscillation Signature (Mukundan et al., 2017)), or event-related potentials (ERP), such as Brain Fingerprinting (BFP) (Farwell, 2009; Farwell & Donchin, 1991; Farwell & Smith, 2001) and the Complex Trial Protocol (CTP) (Rosenfeld et al., 2008). Technologies that use electroencephalographic measures can be categorised under the term *forensic brainwave analysis* (FBA).

BFP is used to detect the presence or absence of concealed knowledge pertaining to an incident in a person's brain by using an extension of the P300 component of the ERP, introduced as P300-MERMER by Farwell and Smith (2001). The P300 is a positive brain potential that is elicited maximally at the mid-line parietal zone (Pz) 300–800 ms post-stimulus when a subject is presented with familiar substantial information embedded within frequent nonsubstantial stimuli (Donchin, 1986; Johnson, 1986; Sutton et al., 1965). The P300 response is not elicited when subjects are presented with stimuli which are not salient to them. The MERMER (memory and encoding related multifaceted electroencephalographic response) was introduced by Farwell and Smith (2001). According to them, the P300-MERMER includes: "(1) the P300, an electrically positive component maximal at the parietal scalp site, (2) another, long latency, electrically negative subcomponent prominent at the frontal scalp site, and (3) phasic changes in the frequency and structure of the signal" (Farwell & Smith, 2001). In the current paper, we refer to P300-MERMER as P300 + LNP epoch – with LNP standing for *late negative potential*. No independent studies have systematically compared the utility of P300 alone versus P300 + LNP epoch. However, Farwell et al. (2013) reported similar accuracy rating of both (i.e., 100% accurate determination) with P300 + LNP epoch resulting in higher statistical confidence on average. In contrast with peak-to-peak (p–p) P300, P300 + LNP epoch uses the full ERP (300–1500 ms) following the stimulus (Farwell, 2012; Farwell et al., 2013; Farwell & Richardson, 2013; Farwell & Richardson, 2022).

BFP compares the ERP of three categories of stimuli: *probes*, *targets*, and *irrelevants*. This method has been referred to as the 3-stimulus protocol (3SP) by Rosenfeld et al. (2017). For crime incidents, probes are produced from the items of information from a crime-scene that should be specifically known to someone who committed, witnessed, or investigated the crime. Probes can be categorised under the personally significant stimuli category, as defined by klein Selle and Ben-Shakhar (2022). Targets are produced from items of information related to the crime and are revealed to subjects before the commencement of the test. In addition, another important role of targets is maintaining subjects' attention during the test by providing a specific behavioural response of pressing a button on the Xbox controller. Irrelevant stimuli are formulated for the purposes of the experiment, are equally plausible as probes and targets, but are unrelated to the incident in question. Hence, targets provide a benchmark baseline P300 + LNP epoch response – for comparison with other ERPs – in all subjects with and without knowledge of the crime. Probe responses will be similar to target responses only in subjects with concealed knowledge of the crime (known as Information Present, or IP). On the other hand, responses to probes by innocent (Information Absent, or IA) subjects should be similar to the response to irrelevant stimuli, as they would not be able to differentiate the two. Since the target information is known to both IP and IA subjects, both elicit P300 + LNP epoch responses to targets (Farwell, 2012; Farwell et al., 2013; Farwell & Donchin, 1991; Farwell & Smith, 2001).

For BFP analysis, the ERPs are compared using a bootstrapped (1000 iterations) double-centered correlation method that results in either an Information-Present classification (IP_C), an Information-Absent classification (IA_C), or Indeterminate. More specifically, BFP calculates and compares the cross-correlation of the ERP signals, between 300 and 1500 ms, for the probe and the target waveforms versus the cross-correlation between the probe and the irrelevant waveforms (done multiple times via bootstrapping). For an IP_C, there must be at least a 90% bootstrapping probability that probes and targets elicited similar ERPs. Mathematically, this would demonstrate that the correlation between the probe and target ERPs is higher than the correlation between the probe and irrelevant ERPs. For an IA_C classification, there must be a 70% or higher bootstrapping probability that the correlation between the probe and irrelevant ERPs is higher than the correlation between the probe and target ERPs. This indicates that probes and irrelevants elicited similar responses and the suspected perpetrator does not possess the intimate knowledge of the crime that comprises probes. If the bootstrapping probability of ERPs does not match that of an IP_C or an IA_C, these subjects are designated Indeterminate, meaning BFP cannot determine whether or not a subject possesses information about the incident in question (Allen & Iacono, 1997; Farwell, 2009, 2012; Farwell et al., 2013; Farwell & Donchin, 1991; Farwell & Smith, 2001; Rosenfeld & Donchin, 2015; Wasserman & Bockenholt, 1989).

All BFP studies conducted by Farwell and his colleagues reported no false negative or false positive determinations (Farwell et al., 2013; Farwell et al., 2014; Farwell & Donchin, 1991; Farwell & Smith, 2001). Farwell and Donchin (1991) showed in a series of two experiments—a mock-espionage scenario (Experiment-1, $n = 20$) and a real-life small legal transgression (Experiment-2, $n = 4$)—that BFP could be used as a guilt detection test in controlled laboratory conditions. All subjects were tested on one IP and one IA incident, resulting in twice as many BFP tests: 40 in Experiment-1 and 8 in Experiment-2. In Experiment-1, five (two IP and three IA) classifications were Indeterminate and the rest were correct classifications of IP_C and IA_C, with no false positives and no false negatives. In Experiment-2, BFP resulted in one Indeterminate (of an IA subject) and the rest were correct IP_C and IA_C classifications (Farwell & Donchin, 1991). Further, Farwell and Smith (2001) tested real-life incidents in six subjects. Each subject was tested with an IP scenario and an IA scenario by a blinded tester. All subjects were correctly classified according to their ground-truth conditions with high bootstrapping probability, no false positives, no false negatives, and no Indeterminates.

Farwell et al. (2013) conducted two further studies examining real-life incidents of felony criminals with no judicial consequences. Their stimuli were developed using interviews, crime-scene investigation, and police and judicial records. Study-1 had 17 IP and 3 IA subjects who were either suspects, convicted criminals claiming innocence, or facing life-imprisonment or capital punishment (Study-2, 9 IP and 5 IA). The ground-truth of subjects was established during post-test interviews by asking subjects to identify the correct probes. Study-1 subjects were correctly classified as per their ground-truth conditions, with no false positives, false negative, or Indeterminates. In Study-2, the IP subjects were offered a US$100,000 prize if they could get BFP to classify them as an IA_C. To produce an IA_C result, they were instructed to use countermeasures such as covert wiggling of a toe within the shoe or imagining being slapped by the experimenter (identical to Rosenfeld et al. (2004) and Mertens and Allen (2008) – more on countermeasures below). These countermeasures were found not effective and all subjects in Study-2 were classified IP_C (Farwell et al., 2013). Notably, none of the previous BFP studies that used P300 + LNP epoch have reported any Indeterminate classifications (Farwell et al., 2013; Farwell et al., 2014; Farwell & Smith, 2001).

Farwell's work has been criticised by other researchers in the field. Meijer et al. (2013), for instance, claimed that some of Farwell's work has not been published in peer-reviewed journals, some subjects' data have been reported twice, the Twenty Scientific Standards did not have a scientific consensus and are applied selectively, that BFP's resistance to countermeasures has not been confirmed by peer-reviewed studies, and Farwell has not provided a clear definition of P300 + LNP epoch. Since we were able to access Farwell's system for independent validation, we came up with a two-goal plan to explore some of the aforementioned criticism. The first goal was to test the accuracy of Farwell's BFP system *as it is* following his protocol and Scientific Standards (20SS hereafter–see Appendix A, also Farwell et al. (2013)) to the letter, to examine whether or not we could produce 100% accurate results. The second goal was to examine BFP's resistance to countermeasures.

1.1 | The first goal – testing the accuracy of BFP

We achieved the first goal – independent examination of the accuracy of BFP – in two studies (Afzali et al., 2022). As planned, we did not modify Farwell's protocol in these studies and made sure that the 20SS were strictly adhered to. In Study-1, life incidents of university students were examined using BFP ($n = 9$ IP and $n = 19$ IA subjects). Although 9 IP and 18 IA subjects were correctly determined by BFP, unlike former studies by Farwell, there was one false positive classification (4.7% error rate). In Study-2, BFP was examined in relation to historical crime incidents of parolees ($n = 3$ IP and $n = 12$ IA subjects). In this instance, BFP correctly determined 2 IP and 6 IA subjects but, again, there was one false positive (8.3% error rate). In addition, Study-2 resulted in 3 Indeterminates (all IA) and 3 (one IP and two IA) additional subjects who could not complete the BFP test. Notably, none of Farwell et al.'s studies reported false positives or subjects who could not complete the BFP test. The Farwell and Donchin (1991) article that reported five Indeterminates used P300 ERPs, rather than P300 + LNP epoch as used in our studies (Afzali et al., 2022). It is worth mentioning that there was a mismatch between the number of IP and IA subjects in our studies. More specifically, there were at least 3–5 IA subjects for each IP subject. Both studies in Afzali et al. (2022) resulted in one false positive determination but no false negatives. A limitation in Afzali et al. (2022), therefore, was the uncertainty regarding the erratic classifications by BFP. We thought it might be that we did not test enough subjects to detect a false negative classification. In addition, it was deemed crucial to examine whether the false positive rate stays one subject per study, as it was in Afzali et al. (2022).

1.2 | The second goal – BFP countermeasures

Considering the limitations in Afzali et al. (2022), we remodified our second goal as a further replication of BFP, as well as testing BFP's resistance to countermeasures. Therefore, a main purpose of Experiment-1 in the current study was to recruit a larger sample matching *n* of IP and IA subjects to explore these issues further. As far as countermeasures are concerned, a few points related to previous research need to be clarified first.

Rosenfeld et al. (2004) conducted two countermeasure studies, both using 3SP – although one of them used six probes and the other used a single probe. The 3SP protocol was designed similar to that of Farwell and Donchin's (1991) BFP experiment. Subjects were divided into three groups: guilty (participated in a mock-crime), innocent (did not participate in the mock-crime in question), and countermeasure (participated in the mock-crime, but guided to engage in specific covert acts). These covert countermeasure acts included wiggling the big toe inside the shoe, or subtly pressing a finger on the leg during the test, or imagining being slapped by the experimenter. According to Rosenfeld et al. (2004), these countermeasures were found to be effective against 3SP either when six probes or when one probe were used. However, the use of countermeasures in one-probe subjects led to longer reaction times. This meant one could detect if subjects were being deceitful (i.e., using countermeasures), leading us to believe that it was more difficult to foil 3SP when one-probe was used than when 6 probes were used. A follow-up study by Mertens and Allen (2008) replicated these findings. Although this study had an overall lesser accuracy of 3SP to begin with, countermeasures led to an even lower guilt detection.

Bergström et al. (2013) conducted a burglary simulation study with the complex Trial Protocol (CTP) guilt detection technique of Rosenfeld et al. (2008) to investigate the potential effectiveness of the Think/No-Think (T/NT) paradigm (Green & Anderson, 2001) as a countermeasure in CTP. CTP was introduced by Rosenfeld et al. (2008) and is an ERP-based FBA system used for guilt detection. T/NT is a cognitive psychology technique used to actively suppress unwanted memories. In the T/NT manipulation, subjects *Think* about one-third of some previously learned information and do not think about another one-third of the information (the *No-Think* condition). The more critical No-Think condition is used to gauge the purposeful suppression of memory. Later, the recall of both 'Think' and 'No-Think' conditions is compared with the remaining one-third baseline items (the previously learned information that is not presented to subjects during the manipulation phase). It has been shown in many studies that the No-Think manipulation results in temporary amnesia of the suppressed information (Anderson & Levy, 2009; Green & Anderson, 2001; Levy & Anderson, 2002, 2012).

Bergström et al. (2013) divided the CTP subjects into three groups: guilty cooperative (participated in the mock-crime and were in the Think condition), guilty uncooperative (participated in the mock-crime and were in the No-Think condition), and baseline (participated in the mock-crime and were not prompted to Think or No-Think). They found that the subjects in the Think condition were classified guilty 22% more than the subjects in the No-Think condition, leading to the conclusion that the No-Think suppression manipulation was an effective countermeasure to CTP. They implied that the No-Think component of the T/NT paradigm could be used as an effective countermeasure against all P300-based FBA measures (Bergström et al., 2013).

In a subsequent study, Hu et al. (2015) used mock-crime scenarios alongside the autobiographical Implicit Association Test (aIAT) to evaluate after-effects of the No-Think suppression manipulation. Two groups of subjects, "guilty" and "suppressed-guilty" committed the mock-crime. Only "suppressed-guilty" were instructed to engage in No-Think suppression during the subsequent CTP test. The third group consisted of "innocent" subjects who did not commit the mock-crime. All subjects participated in CTP followed by the aIAT test. During aIAT, the subjects were asked to classify some factual and mock-crime-related sentences as true or false. The results of the CTP indicated that the "suppressed-guilty" subjects who underwent the No-Think condition were disproportionately misclassified as innocent, consistent with Bergström et al. (2013). Additionally, the aIAT test showed that only subjects in the "suppressed-guilty" condition had delayed responses to the previously supressed stimuli. This demonstrated that not only the recall of unwanted memories was impaired under the No-Think suppression instructions, but so was their automatic influence, as indicated by the delayed access to those memories that had undergone suppression during the aIAT test (Hu et al., 2015).

Collectively, Rosenfeld et al. (2004) and Mertens and Allen (2008) pointed out the susceptibility of 3SP to countermeasures. In addition, Bergström et al. (2013) and Hu et al. (2015) implied that, similar to CTP, all P300-based FBA techniques (including BFP) might be susceptible to cognitive countermeasures. But as Rosenfeld et al. (2017) identified, the conclusions from some of these studies (e.g., Bergström et al. (2013) and Hu et al. (2015)) could not be relied upon due to untested confounding variables (see Rosenfeld et al. (2017) for a review). Rosenfeld et al. (2017) demonstrated that memory suppression could not be used as an effective countermeasure against semantic memories. But it remains unclear if memory suppression might be an effective countermeasure against BFP – which uses real-life autobiographical episodic memories. Among concealed information test (CIT) measures, the CTP has been reported to be resistant to both physical and cognitive countermeasures (Rosenfeld et al., 2008; Rosenfeld et al., 2013; Rosenfeld & Labkovsky, 2010). The question is: could it be that the mock-crime-based scenarios are inherently susceptible to countermeasures due to their artificial nature, but autobiographical episodic memories–which are used in BFP–might be more resistant to countermeasures? The current study is part of a series of studies that explores this possibility. Our lab has designed further studies to examine the rest of the countermeasures used in Rosenfeld et al. (2004) and Mertens and Allen (2008) in the near future.

Another reason for pursuing the current study was Farwell's disagreement with published countermeasure studies. Farwell et al. (2013) contested previously discussed countermeasures findings on the basis that the designs, procedures, and data analysis algorithm used in these studies were critically different from BFP (Farwell et al., 2013; Farwell & Richardson, 2013). More specifically, Farwell et al. (2013) devised 20SS which require specific guidelines to be followed in order for any test to qualify as BFP and asserting that any substantial deviations from this protocol would not be considered 'BFP'. Since these standards were not followed by Bergström et al. (2013) or Rosenfeld et al. (2004), Farwell et al. (2013) stated that their speculations cannot be generalised to BFP. This assertion also appeals to more recent work (Hu et al., 2015; klein Selle et al., 2021) all of which has reasserted claims about the susceptibility of P300-based FBA techniques to countermeasures. It is also worth mentioning that Farwell's paradigm is critically different from that of CTP and similar studies in terms of data analysis. Farwell's BPF relies on bootstrapped cross-correlations of ERP time series between the three types of stimuli. This is referred to as classification algorithm (e.g., Farwell et al., 2013) or bootstrapped correlation difference (e.g., Abootalebi et al., 2006). Rosenfeld's CTP relies on differences between P300 amplitudes (single points in the ERP time series)–referred to as 'bootstrapped amplitude difference' (Abootalebi et al., 2006) or comparison algorithm (Farwell et al., 2013) between probe and irrelevant stimuli. While Farwell and Rosenfeld both claim superiority of their respective paradigms, Abootalebi et al. (2006) reported similar accuracy estimates with both paradigms. Notwithstanding, it should be noted that Abootalebi et al. were only comparing analyses of ERPs up to 900 ms, as opposed to Farwell's P300 + LNP epoch of 300–1500 ms. In addition, another critical difference between CTP and BFP is the definition of "target" stimuli. In BFP, targets are part of the analysis. In other words, the classification is based on the similarity between probe and target stimuli versus probe and irrelevant stimuli. In CTP, on the other hand, the amplitude of probe stimuli is compared with the amplitude of irrelevant stimuli, with no involvement of targets. Lastly, procedure-wise, BFP adheres to a specific set of protocols, 20SS (albeit controversial). However, we could not find any such published procedural guidelines for CTP. Needless to say, procedural integrity is a critical element in experimental design.

One could dismiss the 20SS devised by Farwell due to lack of scientific consensus, as did Meijer et al. (2013), but what if these standards were, in fact, protective against countermeasures? We decided to explore this possibility, especially due to Farwell's work lacking independent replication. Therefore, there has been a gap in literature: some parties claimed that all ERP-based crime detection measures were susceptible to countermeasures, but they have not directly examined BFP relative to these countermeasures. Farwell, on the other hand, refuted these claims but has not yet provided objective empirical evidence to substantiate his claims. It is worth emphasising that CTP has consistently been found to be protected against countermeasures (Rosenfeld et al., 2008; Rosenfeld et al., 2013; Rosenfeld & Labkovsky, 2010). This being the case, the second purpose of the present study (Experiment-2 in the present study) was to focus in more detail on Farwell et al. (2013)'s critique of FBA countermeasure studies (Bergström et al., 2013; Hu et al., 2015; Rosenfeld et al., 2004). Our advantage has been that we had access to the BFP system, were trained by Farwell, and had conducted two replication studies (Afzali et al., 2022). Moreover, the lead researcher of this study (UA) is also a certified tester of the T/NT memory suppression technique that was modified to be used in this study (see below).

In Experiment-2 of the current study, for the first time, we investigated cognitive countermeasures against BFP, while adhering to

Farwell's 20SS. However, there is a procedural clash between No-Think suppression inducement and BFP. The No-Think suppression manipulation requires a subject to directly suppress a response before it enters memory. On the other hand, before each block of BFP, subjects are cued with short descriptions (the incident-related information) corresponding to the critical stimuli. In doing this, the responses that the No-Think instruction is designed to banish from conscious thought by using suppression, may tend to be reactivated with the brief reminder descriptions. Thus, incorporating suppression with BFP contradicts a fundamental procedural constraint inherent in the T/NT paradigm.

To avoid contradiction, the suppression manipulation needs to be modified if it is to be used as a countermeasure against BFP. Hence, rather than asking subjects to suppress a critical stimulus (e.g., probe, target, or irrelevant), they were asked to completely suppress the whole autobiographical incident in question during each BFP test block. In this way, we also avoided the criticism of task demand leading to failure of countermeasures as raised by Rosenfeld in relation to Hu et al. (2015) (see Rosenfeld et al., 2017). Such complete suppression, known as direct-suppression, conforms to the definition of No-Think suppression wherein a subject is asked to stop thinking about a piece of information by directly suppressing it at the moment it tries to enter memory, without replacing it with any other stimulus, thought, or idea (Bergström et al., 2009). Direct-suppression of unwanted memories has successfully been applied to diverse content, including autobiographical memories (Noreen et al., 2016; Noreen & MacLeod, 2013; Stephens et al., 2013). With this as the first step of our countermeasure studies, we will be testing the rest of the countermeasures (Mertens & Allen, 2008; Rosenfeld et al., 2004) in our upcoming endeavours.

We also tested another potential countermeasure known as thought-substitution. During thought-substitution, a subject is instructed to stop thinking about a piece of information by generating alternative thoughts about another scenario, unrelated to the incident in question, to occupy memory (Bergström et al., 2009; Hertel & Calcaterra, 2005; Levy & Anderson, 2002). Hence, in the No-Think direct-suppression condition, subjects suppressed retrieval of an autobiographical memory, whereas in the thought-substitution condition, they were asked to encode a new hypothetical scenario. It is notable that direct-suppression has been reported to be a more effective temporary amnesia-inducing tool than thought-substitution (Bergström et al., 2009).

1.3 | The current study

The current study involves two experiments: Experiment-1 (BFP replication) and Experiment-2 (BFP countermeasures). In Experiment-1, we interviewed 16 subjects and developed corresponding BFP tests for their real-life autobiographical incidents. We also had 16 control subjects to function as ground-truth IA. It was hypothesised that the ground-truth IP subjects would be classified as IP_C and that the ground-truth IA subjects would be classified as IA_C by the BFP system (Hypothesis-1).

In Experiment-2, half of the IP subjects of Experiment-1 (now verified as IP_C) were randomly assigned to direct-suppression and the other half were assigned to thought-substitution conditions, and the BFP test was repeated. In the direct-suppression condition, subjects were instructed to suppress (No-Think) the incident in question when reading the short descriptions before the block and during the whole block. In the thought-substitution condition, subjects were instructed to think about a different unrelated incident when reading the short descriptions before the block and during the block. Extrapolating from Bergström et al. (2013) and Hu et al. (2015), these manipulations would render BFP ineffective (i.e., subjects will be less likely to be correctly classified as IP_C). However, as mentioned earlier, the current BFP deals with real-life incidents and follows the 20SS, while Bergström et al. (2013) and Hu et al. (2015) used fabricated incidents and did not adhere to the 20SS advocated by Farwell et al. (2013). Hence, we hypothesised that despite using direct-suppression or thought-substitution countermeasures, all subjects in Experiment-2 would again be classified as IP_C (Hypothesis-2).

2 | METHOD

2.1 | Participants

A total of 36 subjects, 12 males and 24 females aged from 18 to 52 ($M = 22.7$, $SD = 6.0$), participated in this study. They were students at the University of Canterbury and were recruited through advertisements on social media and posters around the campus. Five subjects were excluded for various reasons, leading to a final $n = 31$. All subjects volunteered to participate. They were informed about the experimental procedure many days in advance and those who agreed to participated, also filled a consent form. Each subject received a shopping mall gift card as gratuity for their participation. The study was approved by the Human Ethics Committee of the University of Canterbury (HEC 2020/12).

2.2 | Material and apparatus

The BFP software and hardware were leased from BFP, LLC (Seattle). Cognionics (San Diego) software was employed to measure electrode impedances on the EEG headset. The experiment was carried out on a Windows PC screen placed at 60 cm in front of a subject. The ERP data were collected with a custom-made, wireless headset that recorded EEG from three midline dry electrodes on the scalp: frontal = Fz, central = Cz, and parietal = Pz (International 10–20 system). Electrooculogram (EOG) signals were used to detect eye-blink artefacts and were collected from Fp1 and Fp2. Linked mastoid electrodes were used as the signal reference. The data from the Pz electrode and EOG signals were amplified with an analogue low-pass filter at 30 Hz and a digital low-pass filter at 6 Hz (3 dB cut-off). Any trials with an EEG > 150 μV and EOG > 200 μV were excluded. The analysis epoch was set at 300–1500 ms from the onset of stimuli based on

TABLE 1 Example of Stimuli

Probe/target	Original stimulus	Description	Irrelevant 1	Irrelevant 2
Probe	Name[a]	Name of a friend who was present at the time	Nadine Cooper	Marlee Morris
Target	Name[a]	Name of a friend who was present at the time	Sophia Gilliam	Charlotte Buckner
Probe	Name[a]	Name of a friend who was present at the time	Nivaan Whitely	Donnie Madden
Target	Name[a]	Name of a friend who was present at the time	Skye Richards	Dave Bishop
Probe	Cigarette pack	Item the subject found after they left the party	Empty wallet	Pocket knife
Target	Ilam	Area where the house was	Shirley	Lincoln
Probe	Horse tranquilliser	Drug one of the friends used	Mushroom tea	Marijuana cookie
Target	Into a bowl	Where the subject vomited	On the couch	Through bed sheets
Probe	Phantom shitter	The nickname given after the incident	Angry druggy	Sexual pest
Target	Excrement on floor	What a parent found in the night	Window was broken	Ripped up carpet
Probe	Started a fight	What the subject almost did at the party	Stole a TV	Smashed a cup
Target	Playground	The place where the subject found something after the party	In gutter	Rooftop

[a]Four names of real people provided by subjects have been redacted to ensure confidentiality.

20SS and previous BFP studies assuming it captures the P300 + LNP epoch. An Xbox controller was used to collect behavioural responses from the subjects.

2.3 | Design

In Experiment-1, the ground-truth of a subject (IP vs. IA) was the between-subjects independent variable and the stimulus type (probe vs. target vs. irrelevant) was the within-subjects independent variable. The ERP response leading to a BFP classification (IP_C, IA_C, or Indeterminate) was the dependent variable. In Experiment-2, the countermeasure condition (direct-suppression vs. thought-substitution) was the between-subjects independent variable and the stimulus type (probe vs. target vs. irrelevant) was the within-subjects independent variable.

For Experiment-1, 16 subjets were randomly chosen and interviewed about a memorable life incident for the ground-truth IP condition. It was ensured that the incident involved them and no one else in the participation pool. A BFP test for each of these incidents was produced (resulting in 16 BFP tests) and the aim was for each test to be used on one ground-truth IP and one ground-truth IA subject.

All interviews and test development were completed by the project leader and a tester. The BFP tests were conducted by two other testers who were not aware of the ground-truth status of their subjects. The testing was strictly monitored by the project leader to ensure that the BFP testing manual and 20SS were adhered to.

2.4 | Stimuli

All subjects were interviewed about a memorable incident of their life cued with when, what, how, where, and why questions so they could recount enough details for the test. They recalled positive, fun, emotional incidents, or when they had a minor brush with the law, as they were assured that the tests and results would be anonymous. The incidents narrated by subjects included an academic trip, overseas and local travel, parties, a police chase, a boating incident, etc. Half of these subjects were randomly selected as ground-truth IP and their incidents were turned into BFP testes as explained below. The rest of subjects, by definition, were ground-truth IA for the selected incidents.

For each incident, a set of stimuli was formed inclusive of 6 probes, 6 targets, and 24 irrelevants according to the 20SS (Farwell et al., 2013). Probes and targets were formed from the information that would be perceived as substantial by a subject who participated or was present during the incident. Irrelevant stimuli were fabricated (two for each probe and two for each target), which were unrelated to the incident in question but equally plausible (see Table 1 for a sample set).

The formulated BFP tests were quite different from each other, as the narrated incidents came from different subjects with different experiences. Therefore, it was particularly important to ensure that the stimuli were developed in an objective manner to avoid methodological biases and flaws. We ensured objective tests were developed by taking the following measures:

1. The stimuli were developed by the project leader and a tester who were both trained and certified by Dr Farwell. Both were experienced BFP test developers through their involvement in the earlier Afzali et al. (2022) studies.
2. The irrelevant items for peoples' names narrated in the interviews were formulated using a database that listed first names and surnames based on their popularity. For instance, if a name narrated by a subject was Emma Anderson, ranking 1 and 11 respectively, their irrelevant was Emily Jackson, ranking 6 and 13, respectively, with a matching number of syllables.
3. The same consistency was observed for names of locations and vehicles.

IP subjects were expected to recognise both probes and targets as pertinent, while IA subjects were expected to only recognise target items. Therefore, IP subjects should produce similar ERP responses (P300 + LNP epoch) to both probes and targets (confirmed by a higher correlation between them), resulting in an IP_C classification. Conversely, IA subjects should demonstrate a P300 + LNP epoch response to targets only. Their probes and irrelevant responses should be similar, resulting in an IA_C classification.

2.5 | Procedure

2.5.1 | Experiment-1: BFP Replication

The interviews were conducted via Zoom and the testing was completed on the University of Canterbury campus. The tests were carried out usually within a week after interviews, and were created in accordance with the 20SS. Each BFP test was made up of 6 probes, 6 targets, and 24 irrelevants. These were divided into two sets (Set 1 and Set 2) each containing 3 probes, 3 targets, and 12 irrelevants (refer to Table 1).

Each stimulus was presented at least 20 times, resulting in 120 probe, 120 target, and 480 irrelevant trials. If any trials were rejected due to artefacts such as blinking or excessive head movement, their replacements were automatically added.

Prior to the BFP test, IA subjects were presented with a list of their targets in order to familiarise themselves and consequently recognise them later during the test to produce their calibration P300 + LNP epoch. IP subjects were presented with a list of their probe and target items, that were narrated to the testers during the interview. This *information confirmation* stage was aimed to ensure that subjects did not lie or overstate the truth, and did not guess any details they did not actually remember. The information confirmation procedure ensured that the incorporated details of the test were verified by the IP subjects. In a field setting during criminal justice studies, this verification could be independently completed with the use of police records, witnesses, and investigators. They then practised a BFP test without ERP recording to learn the experimental procedure. They were instructed to recognise targets and press the left-hand button on the Xbox controller, and to press the right-hand button for other stimuli (probes and irrelevants).

The BFP test

The main task of a subject during a BFP test is to read and recognise a target item and press the left-hand button on the Xbox controller when they do so. Probes, presented to both IP and IA subjects, were to be recognised by the IP subjects only. Irrelevant items were read but required a right-hand button press, since they are unrelated to the incident. To ensure this, a list of irrelevant items was briefly presented, and subjects were asked to identify if any item was crucial to them for any reason. For instance, if an irrelevant item was "Canada" and a subject reported that they were born there, Canada would be changed to another country. Pressing the correct button on the Xbox controller ensured that the subjects were paying attention to the stimuli. Otherwise, this had no relevance to P300 + LNP epoch ERP response generation. However, it is crucial to require such a behavioural response to generate an accurate P300 + LNP epoch representation.

The BFP test was divided into 10 blocks with each block lasting 3–5 min. Sets 1 and 2 of the stimuli were presented in alternating blocks. Each block consisted of at least 72 trials, presenting each stimulus at least four times, with 1/6 probe, 1/6 target, and 4/6 irrelevant trials. The following instructions were presented and read aloud to the subjects before each block: "Here are the items you will see in this test that are related to the investigated situation. Push the left-hand button for the items that were on the short list of things you know about the situation, and the right-hand button for anything else". The *short list* refers to the list of targets that each subject had to recognise and for which a left-hand button press was the appropriate response. With this, a list of three item descriptions for targets and three item descriptions for probes (without being accompanied by the targets, probes, or irrelevants) was presented, such as "Type of car the group travelled in".

A fixation cross (X) was displayed at the centre of the screen for 1000 ms, followed by a stimulus displayed for 300 ms, and a blank screen for 1700 ms followed by another fixation cross signalling the new trial (see Figure 1). Subjects could blink during the fixation cross if they wished to but, otherwise, were asked to inhibit blinks as much as possible and remain immobile. Trials affected by eye blinks, identified by amplitude >200 μV in the Fp1 channel, or by head movement identified by amplitude >150 μV in the Pz channel, were rejected and automatically replaced by new trials. A block continued until a minimum number of artefact-free trials (12 probe, 12 target, and 48 irrelevant) were recorded. At the end of each BFP test, a total of at least 120 probes, 120 targets, and 480 irrelevants are required to complete a BFP classification according to the 20SS. Data were digitised at 100 Hz and electrode-scalp impedances did not exceed 10 kΩ. After 10 blocks were completed, the IP subjects were requested to attend the next experiment (Experiment-2) at a designated date and time. The IA subjects received a gratuity and were debriefed.

2.5.2 | Experiment-2: BFP countermeasures

Experiment-2 was conducted on the 16 IP subjects of Experiment-1. It was similar to Experiment-1 except: Prior to presenting and reading aloud the block instructions, the eight subjects in the direct-suppression group were verbally instructed to not think about the incident in question while looking at and hearing the descriptions, and during the block. It was emphasised that they should try their best to pay attention to the stimuli on the screen, and press the corresponding correct button on the Xbox controller, while actively suppressing the event in question when they read probes, targets, or irrelevants, but not by replacing the event in question with any other incident or memory. The rest of the experiment was similar to Experiment-1.

For the eight subjects in the thought-substitution group, the tester elicited and discussed with them another memorable incident (Event 2) in the subject's life and this was noted. Event 2 was given a

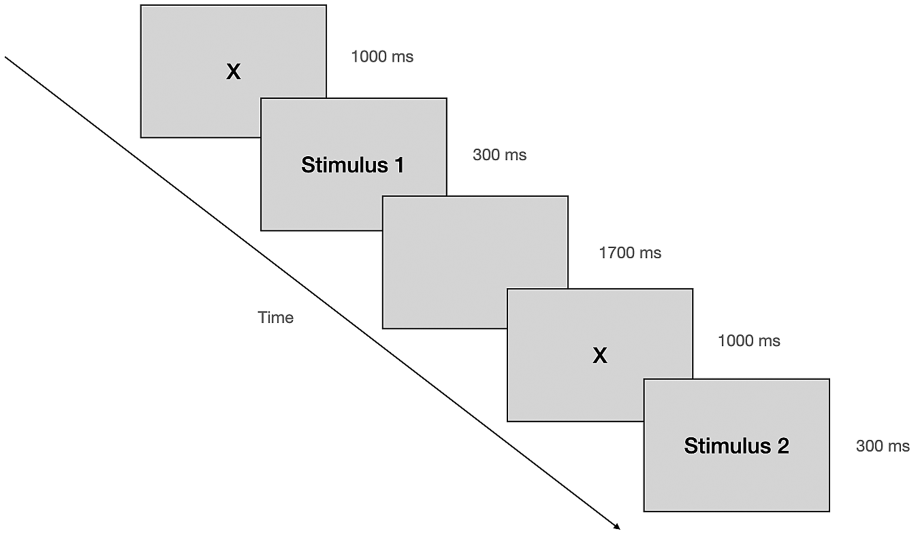

FIGURE 1 Paradigm figure of the Brain Fingerprinting experimental procedure

specific name that ensured that the subject recognised it when referred to during the experiment. More importantly, it was made sure that the details of Event 2 did not overlap with Event 1. Prior to presenting and reading aloud the block instructions, the subjects were verbally instructed to substitute thinking of Event 1 (i.e., the name of their original incident in Experiment-1) to thinking about Event 2 (i. e., the alternative incident) while looking at and hearing the descriptions, and during the block. It was emphasised that they should try to complete the experiment exactly as before while actively substituting thinking about Event 1 to thinking about Event 2 during the task. Following this, the block of stimuli was presented. After the experiment was concluded, the subjects received their vouchers and were debriefed.

2.6 | Data analysis

BFP uses bootstrapping procedures (Farwell et al., 2013; Wasserman & Bockenholt, 1989) and double-centered correlations to decide whether the probe ERPs are more correlated with target ERPs (large P300 + LNP epoch amplitudes for both, resulting in an IP_C classification) or irrelevant ERPs (lacking a large P300 amplitude for probe resulting in an IA_C classification). The P300 + LNP epoch is defined as the ERP waveform 300–1500 ms post-stimulus, similar with previous BFP studies by Farwell (e.g., Farwell et al., 2013). Here, BFP calculates and compares the cross-correlation of the ERP signals for the probe and the target waveforms versus the cross-correlation between the probe and the irrelevant waveforms. The determination decision is made based on a bootstrapping probability wherein for each subject's P300 + LNP epoch ERPs, P probe responses, T target responses, and I irrelevant responses are randomly sub-sampled such that P equals the number of probe trials in the data set, T equals the number of target trials in the data set, and I equals the number of irrelevant trials in the data set.

BFP compares the time-series correlation between the response curves for probes and targets with the correlation between the response curves for probes and irrelevants. After repeating this 1000 times, the frequency of the probe-target correlation being greater than the probe-irrelevant correlation is converted to a percentage. We call this percentage bootstrapping probability of being IP_C (possessing concealed knowledge about the incident). It is interpreted as the bootstrap probability that a subject possesses concealed knowledge about the incident that they are tested on (IP_C). The a priori cut-off for an IP_C is set at 90% (Farwell, 2009; Farwell et al., 2013). This being the case, if the resultant percentage for Subject A is 95 (for instance), their probability of being classified as IP_C would be 95%. The a priori cut-off for an IA_C classification (not possessing the concealed knowledge) is set at 70% in the opposite direction of the IP_C determination (100 minus IP_C Bootstrapping probability) (Farwell et al., 2013; Farwell et al., 2014). With this being the case, Subject A's probability of being IA_C is (100–95 = 5), or 5%. Therefore, Subject A is classified as an IP_C with a 95% bootstrapping probability because 95 is above 90% threshold for IP_C classification, but 5 is below the 70% threshold for IA_C classification.

As another example, if Subject B has a percentage of 15, by definition, their probability of being classified as IP_C would be 15%, and of being classified as IA_C would be 85% (100–15 = 85). Therefore, Subject B would be classified as IA_C with an 85% bootstrapping probability since their IP_C probability is below the 90% (a priori threshold for the IP_C classification), but their IA_C probability is above the a priori threshold of 70% for an IA_C classification. Any subjects who do not meet either of these criteria are categorised as Indeterminate (Farwell & Donchin, 1991).

Subjects were required to overtly press a button on the Xbox controller to ensure that they understood and paid attention to each stimulus. The frequency of correct behavioural responses (pressing the left button for targets and the right button for probes and irrelevants) is calculated as a percentage score by the BFP software. We set an a priori criterion of ≥80% behavioural accuracy for each block for target and irrelevant stimuli. The below perfect (100%) standard was decided to cater for unintentional pressing of a wrong button

TABLE 2 Summary results of 16 scenarios in Experiment-1 (BFP replication)

Subject ID	Incident	Ground-truth	BFP determination	Bootstrapping probability (%)
S01	School contest	IP	IP_C	99.0
S02		IA	IA_C	98.4
S03	Canada trip	IP	IP_C	99.9
S04		IA	IA_C	99.9
S05	Insect repellent	IP	IP_C	97.0
S06		IA	IA_C	78.3
S07	Sustainability prize	IP	IP_C	99.9
S08		IA	IA_C	99.9
S09	Police car	IP	IP_C	99.9
S10		IA	IA_C	77.3
S11	Bush fire	IP	IP_C	99.9
S12		IA	Withdrew due to eye fatigue	
S13		IA	IA_C	98.8
S14	Sea witch	IP	IP_C	99.9
S15		IA	IA_C	99.9
S16	Trip to queenstown	IP	IP_C	99.6
S17		IA	IA_C	98.6
S18	Street signs	IP	IP_C	99.9
S19		IA	IA_C	99.6
S20	12 pubs of xmas	IP	IP_C	96.3
S21		IA	IA_C	99.4
S22	Motion sickness	IP	IP_C	99.9
S23		IA	IA_C	90.7
S24	Horse riding	IP	IND[a]	88.3
S25		IA	IA_C	97.6
S26	House party	IP	IP_C	99.3
S27		IA	Withdrew due to eye fatigue	
S28		IA	Withdrew due to eye fatigue	
S29	Bad mosquitoes	IP	IP_C	93.7
S30		IA	IP_C[b]	98.5
S31	Representing UC	IP	IP_C	96.4
S32		IA	Data loss	
S33		IA	IA_C	99.7
S34	Trip to Vietnam	IP	IP_C	92.8
S35		IA	Invalid test	
S36		IA	IA_C	86.9

[a]Blue coloured font shows an indeterminate classification.
[b]Red coloured font shows a false positive classification.
Abbreviations: BFP, brain fingerprinting; IP, ground-truth information present; IP_C, classified as information present by brain fingerprinting; IA, ground-truth information absent; IA_C, classified as information absent by brain fingerprinting; IND, classified as indeterminate by brain fingerprinting.

based on our prior experience with BFP testing. A block would be rejected if the accuracy dropped below 80%. In such a condition, an extra block would be completed to reach the total required number of trials as stipulated in the 20SS. To prevent possible tester bias, scrutiny of these accuracies and deciding whether more blocks would need to be run were taken by the project leader.

3 | RESULTS

3.1 | Exclusions

One subject's (S32) data were lost due to a software malfunction. Two subjects (S12 and S27) withdrew during Experiment-1 due to

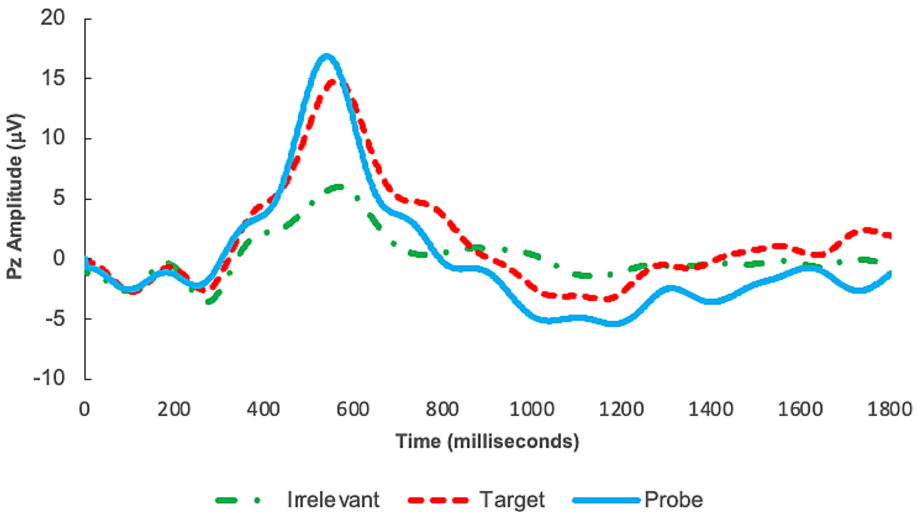

FIGURE 2 BFP Response Waveforms of S03 in "Canada Trip", Experiment-1 (IP → IP_C). BFP, Brain Fingerprinting; IP, ground-truth information present; IP_C, classified as information present by brain

uncontrollable excessive eye-blinking leading to eye fatigue. Three new replacement subjects, S33, S13, and S28 were recruited. However, S28 also had excessive eye-blinking and was excluded. Another subject's (S35) test was determined to be invalid as their behavioural accuracy did not reach the minimum requirement of 80%. Upon further enquiry after the test, she reported being pre-occupied with an upcoming job interview. This subject was then replaced by S36. Although the aim was to replace any excluded subjects, S28 was unable to be replaced due to the Covid-19 pandemic restrictions and the available time frame. This resulted in 31 subjects, rather than the required 32 for 16 tested incidents (see Table 2 for details). All remaining subjects met the behavioural accuracy criterion (≥ 80%).

3.2 | Experiment-1 (BFP replication) findings

As displayed in Table 2, of the 16 IP subjects, 15 were correctly classified as IP_C, with a mean bootstrapping probability of 98.2% (*Med* = 99.6%). Since the data were skewed, we used the non-parametric alternative of the single sample *t*–test (Wilcoxon rank test) to test the hypothesis. The test revealed that this median was significantly higher than the IP_C threshold, $W = 120$, $p < .001$, with the lower bound of CIs higher than 90 (95% CI: [92.8, Inf]). Assuming a binomial distribution (e.g., x number successes out of n trials), we conducted a binomial test using the binom.test() in *R* (R Core Team, 2022). This *R* package also produces confidence interval limits (or fiducial limits) as illustrated by Clopper and Pearson (1934). For 15 out of 16 successes (IP_C determinations), the probability of success is 0.94 [0.70, 1.00]. Simulations showed that any large sample approximation with continuity correction gives almost the same result. One subject (S24) was Indeterminate, with a bootstrapping probability of 88.3%.

Of the 15 IA subjects, 14 were correctly classified as IA_C, with a mean bootstrapping probability of 94.6% (*Med* = 98.7%). A single sample Wilcoxon test showed that this median was significantly higher than the IA_C threshold, $W = 105$, $p < .001$, with the lower bound of CIs higher than 70 (95% CI: [79.7, Inf]). The binomial test with confidence intervals (Clopper & Pearson, 1934) for 14 out of 15 successes (IA_C determination) showed a success probability of 0.93 [0.68, 1.00]. One IA subject (S30) was misclassified as IP_C with a bootstrapping probability of 98.5%. As a result, the classification accuracy was 96.8% with the false positive rate of 3.2%. See Figures 2–5 for example ERPs for an IP_C (S03), an IA_C (S04), the Indeterminate S24, and the false-positive S30, respectively. The rest of the ERPs are provided in Appendix B.

3.3 | Experiment-2 (BFP countermeasures) findings

All IP subjects in Experiment-1 were randomly assigned to either direct-suppression or thought-substitution conditions and were re-tested. As Table 3 shows, none of these manipulations were successful at decreasing the BFP bootstrapping probability to less than 90%, which would otherwise be an Indeterminate or a false negative classification. Therefore, we notice 16 IP_C determinations. Notably, the previously Indeterminate subject (S24) has now been classified as IP_C (consistent with the ground-truth). The average bootstrapping probability of IP_C determinations was 98.9% (including the re-testing of S24 who was Indeterminate in Experiment-1). We also conducted a one-tailed paired samples Wilcoxon rank test between the bootstrapping probabilities of 15 subjects who were correctly classified as IP_C in Experiment-1 and re-tested in Experiment-2 as result of either manipulations. The median bootstrapping probability was 99.9% (cf. *Med* = 99.6% for IP_C in Experiment-1; and only a decrease from *Med* = 99.6 could indicate a successful countermeasure). The difference between bootstrapping probabilities in Experiment-1 and Experiment-2 was not significant, $W = 18.0$, $p = .846$. With the Clopper and Pearson (1934) method, the probability of success (an IP_C result in Experiment-2) for 8 out of 8 trials was 1.00 [0.63, 1.00]. Based on the Jeffrey's interval method (Brown et al., 2001; Dean & Pagano, 2015), the probability of success was >0.79 at $\alpha = .05$ with

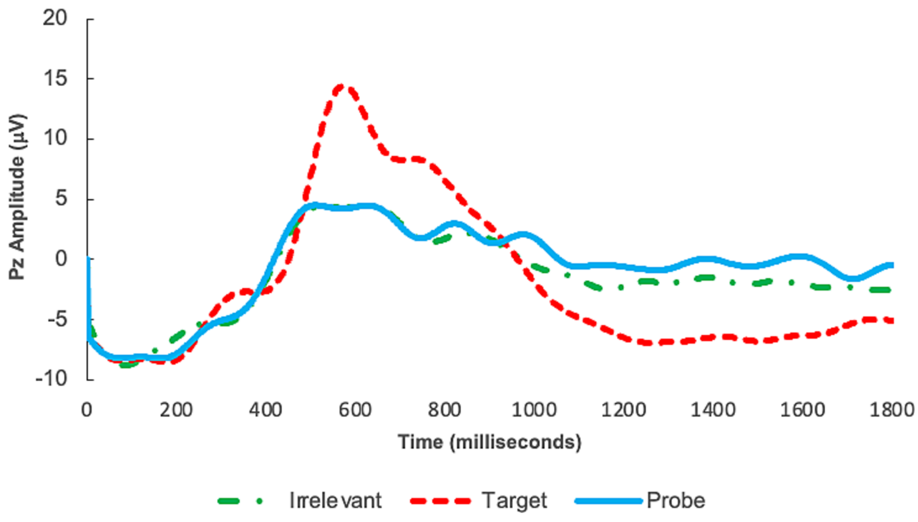

FIGURE 3 BFP Response Waveforms of S04 in "Canada Trip" Experiment-1 (IA → IA_C). BFP, Brain Fingerprinting; IP, ground-truth information present; IA, ground-truth information absent; IA_C, classified as information absent by brain fingerprinting

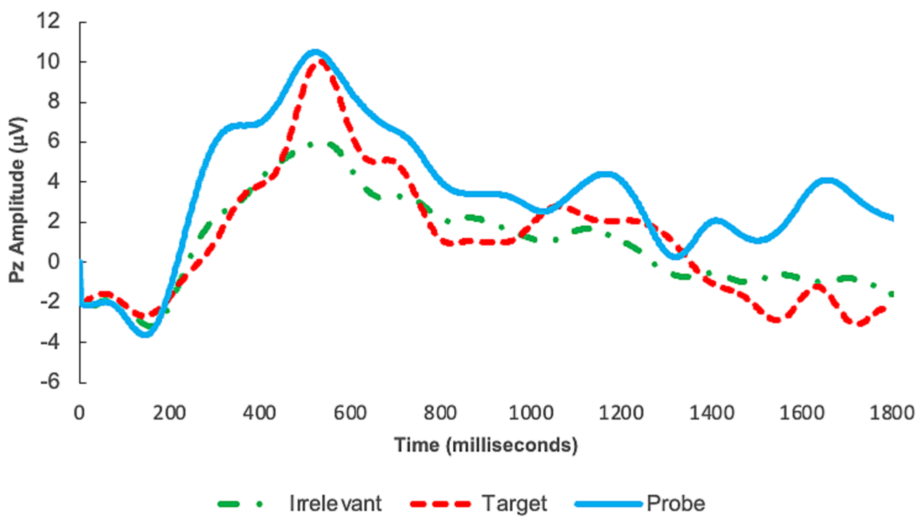

FIGURE 4 BFP Response Waveforms of S24 in "Horse Riding" Experiment-1 (IP → IND). BFP, Brain Fingerprinting; IP_C, classified as information present by brain fingerprinting; IND, classified as indeterminate by brain fingerprinting.

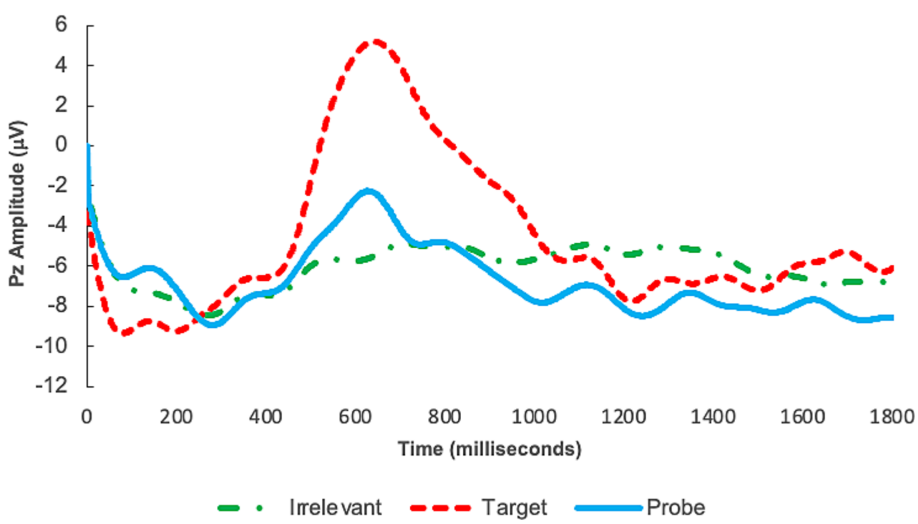

FIGURE 5 BFP Response Waveforms of S30 in "Bad Mosquitoes" Experiment-1 (IA → IP_C). BFP, Brain Fingerprinting; IP_C, classified as information present by brain fingerprinting; IND, classified as indeterminate by brain fingerprinting.

TABLE 3 Summary results Experiment-2 (BFP countermeasures) versus Experiment-1 (BFP replication)

Subject ID	Incident	Experiments	BFP determination	Bootstrapping probability (%)	Condition
S01	School contest	Experiment-1	IP_C	99.0	
		Experiment-2	IP_C	99.5	Substitution
S03	Canada trip	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.9	Substitution
S05	Insect repellent	Experiment-1	IP_C	97.0	
		Experiment-2	IP_C	94.5	Suppression
S07	Sustainability prize	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.9	Suppression
S09	Police car	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.8	Substitution
S11	Bush fire	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.5	Suppression
S14	Sea witch	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.9	Suppression
S16	Trip to queenstown	Experiment-1	IP_C	99.6	
		Experiment-2	IP_C	99.9	Suppression
S18	Street signs	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.9	Suppression
S20	12 Pubs of Xmas	Experiment-1	IP_C	96.3	
		Experiment-2	IP_C	99.8	Substitution
S22	Motion sickness	Experiment-1	IP_C	99.9	
		Experiment-2	IP_C	99.9	Substitution
S24	Horse riding	Experiment-1	IND[a]	88.3	
		Experiment-2	IP_C	96.7	Substitution
S26	House party	Experiment-1	IP_C	99.3	
		Experiment-2	IP_C	99.9	Substitution
S29	Bad mosquitoes	Experiment-1	IP_C	93.7	
		Experiment-2	IP_C	99.9	Suppression
S31	Representing UC	Experiment-1	IP_C	96.4	
		Experiment-2	IP_C	92.8	Suppression
S34	Trip to Vietnam	Experiment-1	IP_C	92.8	
		Experiment-2	IP_C	99.9	Substitution

[a]Blue coloured font shows an indeterminate classification.
Abbreviations: BFP, brain fingerprinting; IP_C, classified as information present by brain fingerprinting; IND, classified as indeterminate by brain fingerprinting.

95% CI: [0.79, 1.00]. As detailed in Table 3, memory-suppression and thought-substitution, each, had 8 IP_C determinations. Consequently, both binomial tests in Experiment-2 and the previous ones in Experiment-1 indicate a perfect or near perfect success with little chance that we might be at around 70% - 80% classification accuracy. Therefore, Hypothesis-2 is supported confirming that the present countermeasures of memory-suppression and thought-substitution are ineffective countermeasures for BFP. Figures 6 and 7 show examples of the BFP waveforms for S03 and S24 in Experiment-2 (See Appendix B for the rest of the ERPs).

4 | DISCUSSION

This study aimed to examine the classification accuracy of BFP and the effects of direct-suppression and thought-substitution countermeasures on the standard BFP test. It is also the first study to have employed a modification of the T/NT paradigm with BFP. This was our third independent examination of the BFP forensic system and technique.

In Experiment-1, the BFP system correctly classified 15 IP subjects and 14 IA subjects, with a false positive and one IP resulting in

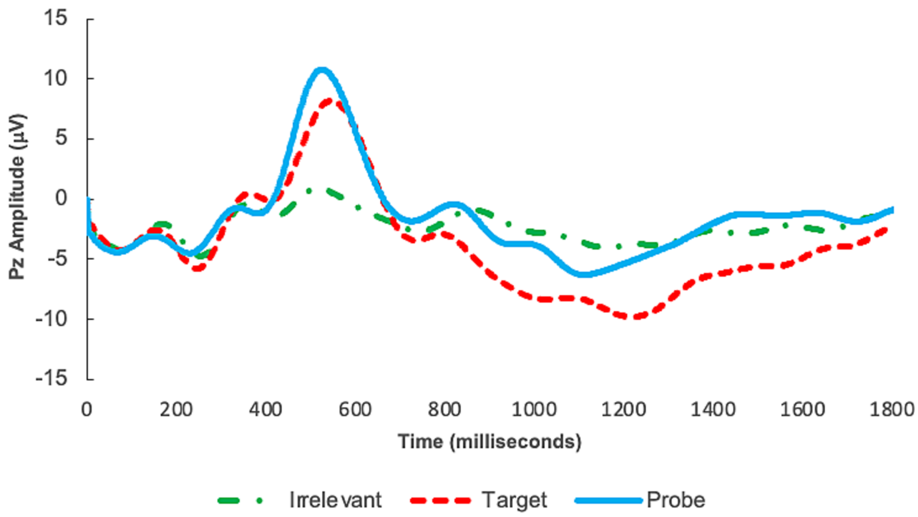

FIGURE 6 BFP Response Waveforms of S03 in "Canada Trip", Experiment-2 (IP → IP_C → IP_C). BFP, Brain Fingerprinting; IP_C, classified as information present by brain fingerprinting; IND, classified as indeterminate by brain fingerprinting.

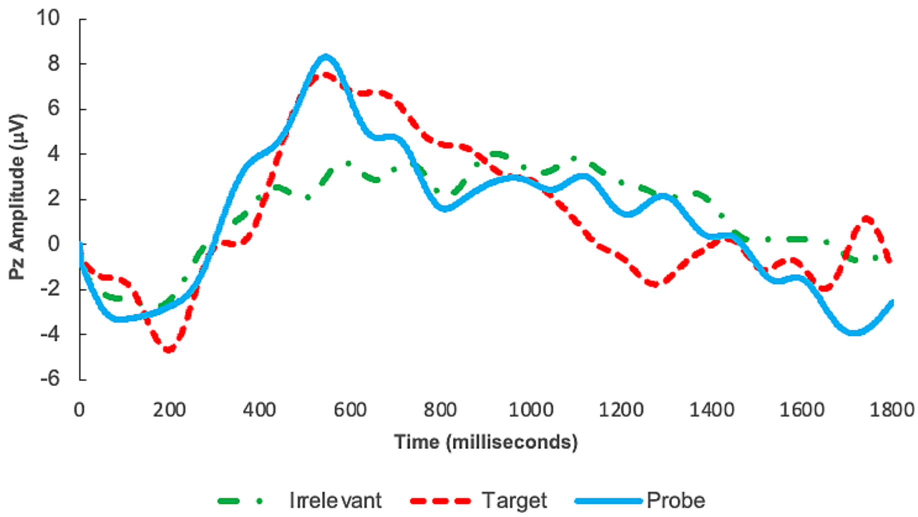

FIGURE 7 BFP Response Waveforms of S24 in "Horse Riding", Experiment-2 (IP → IND → IP_C). BFP, Brain Fingerprinting; IP, ground-truth information-absent; IP_C, classified as information present by brain fingerprinting; IND, classified as indeterminate by brain fingerprinting

an Indeterminate designation. Despite having a high classification accuracy, BFP did not produce 100% accurate classifications as claimed in previous studies (Farwell, 2009, 2012; Farwell et al., 2013; Farwell & Makeig, 2019; Farwell & Smith, 2001); therefore, Hypothesis-1 was not supported. These findings are, however, consistent with our previous findings, Afzali et al. (2022), that reported one false positive in each of two BFP studies (with 4.7% and 8.3% false positive rates, respectively). Nonetheless, all three independent studies by our team at University of Canterbury resulted in one false-positive classification each (average error rate of three studies = 5.4%). The current findings indicate that despite an increased IP sample, there were no more false-positives. In addition, we did not detect any false-negative classifications.

In Experiment-2, neither direct-suppression nor thought-substitution was effective as a countermeasure against BFP. In addition, the previously Indeterminate (S24) was correctly classified as IP_C in Experiment-2, despite the thought-substitution manipulation. These findings confirm Hypothesis-2. In addition, one might speculate that the distinction between episodic memories and semantic memories, as considered by Rosenfeld et al. (2017), might explain the resistance of BFP to these cognitive countermeasures. Nonetheless, it is worth noting that unless BFP is tested against the specific countermeasures that Rosenfeld et al. (2008) used against CTP, it is not possible to claim that BFP is superior to other P300-based CIT measures in terms of robustness to countermeasures.

It should be noted that the false-positive subject in Experiment-1 (S30) was compliant, had followed instructions, met all BFP Scientific Standards, and showed a high level of behavioural accuracy. The tester confirmed that the subject had not been made aware of the incident in question

On the other hand, the Indeterminate (S24) reported that they were sensitive to the blue colour (the screen background colour during BFP trials). However, since they were correctly classified as IP_C in Experiment-2, this probably indicates that subjects had become skilled as they were performing the BFP task for the second time in Experiment-2, which might have competed with the countermeasure manipulation. This is further supported by multiple reports from Experiment-2 subjects who found the task relatively easier and more effortless.

Importantly, three subjects of Experiment-1 withdrew due to excessive eye-blinking and inability to complete the BFP test. Such an outcome has not been reported in previous BFP studies by Farwell. However, we reported three such subjects (parolees) in our recent BFP study too (Afzali et al., 2022). These findings indicate that BFP might not be a suitable concealed knowledge detection test for everyone, for there are people who cannot complete it.

Participating in 10 extra blocks due to participating in Experiment-1 could be considered a possible limitation due to prior practice or the probes having been imprinted more deeply into memory. A future study to replicate Experiment-2 (with countermeasures) without subjects having participated in a prior Experiment-1 would shed light on this possibility.

The choice of subjects is also a limitation in this study. BFP is designed to detect concealed knowledge in criminals. However, conclusions of this study are based on university students. It is possible that real-life criminals might be able to use memory-suppression or thought-substitution strategies to render BFP ineffective. Nonetheless, unless a future study confirms this, such a claim cannot be made with any confidence.

Another limitation of this study has been the use of the extra information confirmation procedure. Though we deemed it needed for this study, it might negatively affect the generalisability and ecological validity of this study. Therefore, we have designed and preregistered a follow-up study to systematically compare subjects with and without using information confirmation procedure to figure out if the determination accuracy is different. Unless this systematic comparison is conducted, we recommend that further BFP studies should not be conducted using information confirmation.

There are some other distinctions to be made between BFP and other FBA studies. Except for Farwell and Donchin (1991), all subsequent BFP studies developed critical stimuli from real-life incidents. By contrast, Bergström et al. (2013), Hu et al. (2015), and Rosenfeld et al. (2008) used mock-crime simulated scenarios in their FBA endeavours. As is stands, the incidents in questions are *real* and autobiographical for BFP studies and *fabricated* for other P300-based FBA methods. In addition, as elaborated in the Method section, before each block of the BFP test, subjects are reminded of the incident by way of *short descriptions* of the upcoming stimuli during the block. These short descriptions play a crucial role, for if a subject knows the incident by way of participating in it, these *short descriptions* remind them of that incident and they elicit P300 + LNP epoch responses to the probe stimuli, as well as the target stimuli, resulting in an IP_C categorisation. On the other hand, since an IA subject does not know the incident, the short descriptions only remind them of the target stimuli that they know from common knowledge or have explicitly been made aware of. Consequently, they elicit P300 + LNP epoch responses to the target stimuli only, resulting in an IA_C categorisation. This difference could potentially explain the failed countermeasure endeavour in the current study. It could be that the No-Think manipulation only works well when the information is newly learned (as it is in the classic T/NT task), and not when the information is more personal and based on real-life situations. This would effectively mean that the countermeasures would not work no matter how much effort a subject applied in suppressing the incident in question or substituting it with a different incident.

As Meijer et al. (2013) noted, there is no scientific consensus on 20SS, and they are considered Farwell's subjective views. Since our three studies (Afzali et al., 2022 and the current study) resulted in >90% determination accuracy, while other studies (e.g., a meta-analysis by Meijer et al., 2014), showed similar accuracy without following any such standards, it is important to scientifically examine these standards and determine which of the 20 are actually crucial to achieving BFP's high determination accuracy. Last, but not least, an important future direction is to examine the accuracy of BFP to countermeasures used by Rosenfeld et al. (2004) and Mertens and Allen (2008). These studies have been planned and will be conducted in the near future.

The current study indicates that BFP is resistant to direct-suppression and thought-substitution countermeasures – noting that these findings should not be generalised to any other countermeasures beyond memory suppression. These cognitive manipulations were not effective in concealing the information that BFP successfully revealed. In addition, the current study, coupled with Afzali et al. (2022), shows that BFP has a high classification accuracy, similar to other FBA measures (e.g., CTP), and that the accuracy is not 100%, as reported in Farwell's publications. Notwithstanding, this is not considered a major flaw. Iacono (2008) pointed out that many other forensic tests have a false positive rate of 2–5%. Even if administered properly, psychophysiological memory detection techniques can result in false-positive outcomes. Nevertheless, the scientific community values ERP-based guilt detection tests more highly than autonomic guilt detection measures (Iacono, 2008; Meijer et al., 2013).

In conclusion, we have presented preliminary evidence that BFP can almost be successfully replicated with a student sample by demonstrating a relatively high classification accuracy, but not as robust as the findings reported in the previous BFP studies by Farwell and colleagues. Moreover, we have also presented preliminary evidence that the modified direct-suppression and thought-substitution manipulations could not diminish the accuracy determination of BFP. Collectively, these findings indicate that, after its susceptibility to countermeasures is tested in persons with criminal histories, BFP has considerable potential to replace autonomic and other conventional crime detection techniques and provides an important addition to the armamentarium of current forensic technologies. Importantly, future studies must be conducted to address the discussed limitations and limited generalisability of the current findings.

AUTHOR CONTRIBUTIONS

M. Usman Afzali: Conceptualization (lead), Methodology (lead), Formal analysis (lead), Investigation (lead), Resources (lead), Data Curation (lead), Writing - Original Draft (lead), Writing - Review & Editing (equal), Visualisation (lead). Richard D. Jones: Conceptualization (supporting), Methodology (supporting), Writing - Review & Editing (equal). Alex P. Seren-Grace: Investigation (supporting), Resources (equal), Writing - Review & Editing (equal). Robin W. Palmer: Writing - Review & Editing (equal). Dena Makarious: Investigation (equal), Data

Curation (equal), Writing - Review & Editing (equal). Mariana N. B. Rodrigues: Investigation (equal), Data Curation (equal), Writing - Review & Editing (equal). Ewald Neumann: Conceptualization (supporting), Methodology (supporting), Writing - Review & Editing (equal).

ACKNOWLEDGMENT

Open access publishing facilitated by University of Canterbury, as part of the Wiley - University of Canterbury agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

We have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

Data subject to third party restrictions. The raw data of Brain Fingerprinting cannot be exported outside the system that is copyright of Dr Farwell (the inventor of the system)

ORCID

M. Usman Afzali https://orcid.org/0000-0001-5119-9388
Richard D. Jones https://orcid.org/0000-0003-2287-3358
Alex P. Seren-Grace https://orcid.org/0000-0003-3084-1452
Robin W. Palmer https://orcid.org/0000-0003-1949-4146
Mariana N. B. Rodrigues https://orcid.org/0000-0001-8387-2012
Ewald Neumann https://orcid.org/0000-0001-6494-0294

REFERENCES

Abootalebi, V., Moradi, M. H., & Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *International Journal of Psychophysiology*, *62*(2), 309–320. https://doi.org/10.1016/j.ijpsycho.2006.05.009

Afzali, M. U., Seren-Grace, A. P., Palmer, R. W., Neumann, E., Makarious, S., Wilson, D., & Jones, R. D. (2022). Detection of concealed knowledge via the ERP-based technique Brain Fingerprinting: Real-life and real-crime incidents. *Psychophysiology*, *59*(11), 1–21. https://doi.org/10.1111/psyp.14110.

Allen, J. J. B., & Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology*, *34*(2), 234–240. https://doi.org/10.1111/j.1469-8986.1997.tb02137.x

American Psychological Association. (2004). The truth about lie detectors (aka polygraph tests).

Anderson, M. C., & Levy, B. J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science*, *18*(4), 189–194. https://doi.org/10.1038/35066572

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology*, *94*(1), 1–11. https://doi.org/10.1016/j.biopsycho.2013.04.012

Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, *48*(4), 726–737. https://doi.org/10.1016/j.neuroimage.2009.06.051

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*(2), 101–133. https://doi.org/10.1214/ss/1009213286

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404–413. https://doi.org/10.2307/2331986

Dean, N., & Pagano, M. (2015). Evaluating confidence interval methods for binomial proportions in clustered surveys. *Journal of Survey Statistics and Methodology*, *3*(4), 484–503. https://doi.org/10.1093/jssam/smv024

Donchin, B., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (1986). Cognitive psychophysiology and human information processing. In E. Donchin & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 244–267). Guilford Press.

Farwell, L. A. (2009). Brain fingerprinting: Detection of concealed information. *Wiley Encyclopedia of Forensic Science*, 1–12. https://doi.org/10.1002/9780470061589.fsa1013

Farwell, L. A. (2012). Brain fingerprinting: A comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics*, *6*(2), 115–154. https://doi.org/10.1007/s11571-012-9192-2

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, *28*(5), 531–547. https://doi.org/10.1111/j.1469-8986.1991.tb01990.x

Farwell, L. A., & Makeig, T. H. (2019). Farwell brain fingerprinting in the case of Harrington v. *Stat*, 7–10. http://www.larryfarwell.com/pdf/OpenCourtFarwellMakeig-dr-larry-farwell-brain-fingerprinting-dr-lawrence-farwell.pdf

Farwell, L. A., & Richardson, D. C. (2013). Brain fingerprinting: Let's focus on the science - a reply to Meijer, Ben-Shakhar, Verschuere, and Donchin. *Cognitive Neurodynamics*, *7*(2), 159–166. https://doi.org/10.1007/s11571-012-9238-5

Farwell, L. A., Richardson, D. C., & Richardson, G. M. (2013). Brain fingerprinting field studies comparing P300-MERMER and P300 brainwave responses in the detection of concealed information. *Cognitive Neurodynamics*, *7*(4), 263–299. https://doi.org/10.1007/s11571-012-9230-0

Farwell, L. A., Richardson, D. C., Richardson, G. M., & Furedy, J. J. (2014). Brain fingerprinting classification concealed information test detects US navy military medical information with P300. *Frontiers in Neuroscience*, *8*, 1–21. https://doi.org/10.3389/fnins.2014.00410

Farwell, L. A., & Richardson, G. M. (2022). Brain fingerprinting field study on major, terrorist crimes supports the brain fingerprinting scientific standards hypothesis: Classification concealed information test with P300 and P300-MERMER succeeds; comparison CIT fails. *Cognitive Neurodynamics*, *17*, 63–104. https://doi.org/10.1007/s11571-022-09795-1

Farwell, L. A., & Smith, S. S. (2001). Using brain MERMER testing to detect knowledge despite efforts to conceal. *Journal of Forensic Science*, *46*(1), 135–143. https://doi.org/10.1520/JFS14925J

Green, C., & Anderson, M. C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369. https://doi.org/10.1038/35066572

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review*, *12*(3), 484–489. https://doi.org/10.3758/BF03193792

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science*, *26*(7), 1098–1106. https://doi.org/10.1177/0956797615575734

Iacono, W. G. (2008). The forensic application of "brain fingerprinting:" why scientists should encourage the use of P300 memory detection methods. *The American Journal of Bioethics*, *8*(1), 30–32. https://doi.org/10.1080/15265160701828550

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*(4), 367–384. https://doi.org/10.1111/j.1469-8986.1986.tb00649.x

Klein Selle, N., & Ben-Shakhar, G. (2022). A new theoretical perspective on concealed information detection. *Psychophysiology*, *e14187*, e14187. https://doi.org/10.1111/psyp.14187

klein Selle, N., Waxman, D., Volz, K., Ambach, W., & Ben-Shakhar, G. (2021). Is the CIT susceptible to misleading information? A constructive replication. *Journal of Forensic Sciences*, *66*(2), 646–655. https://doi.org/10.1111/1556-4029.14630

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, *6*(7), 299–305. https://doi.org/10.1016/S1364-6613(02)01923-X

Levy, B. J., & Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. *The Journal of Neuroscience*, *32*(47), 16785–16794. https://doi.org/10.1523/jneurosci.2640-12.2012

Meijer, E. H., Ben-Shakhar, G., Verschuere, B., & Donchin, E. (2013). A comment on Farwell (2012): Brain fingerprinting: A comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics*, *7*, 155–158. https://doi.org/10.1007/s11571-012-9217-x

Meijer, E. H., Klein Selle, N., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the C oncealed I nformation T Est: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*(9), 879–904. https://doi.org/10.1111/psyp.12239

Mertens, R., & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, *45*(2), 286–298. https://doi.org/10.1111/j.1469-8986.2007.00615.x

Mukundan, C., Sumit, S., & Chetan, S. (2017). Brain electrical oscillations signature profiling (BEOS) for measuring the process of remembrance. *EC Neurology*, *8*(6), 217–230.

Noreen, S., & MacLeod, M. D. (2013). It's all in the detail: Intentional forgetting of autobiographical memories using the autobiographical think/no-think task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 375–393. https://doi.org/10.1037/a0028888

Noreen, S., O'Connor, A. R., & MacLeod, M. D. (2016). Neural correlates of direct and indirect suppression of autobiographical memories. *Frontiers in Psychology*, *7*, 1–18. https://doi.org/10.3389/fpsyg.2016.00379

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.r-project.org

Rosenfeld, J. P., & Donchin, E. (2015). Resampling (bootstrapping) the mean: A definite do. *Psychophysiology*, *52*(7), 969–972. https://doi.org/10.1111/psyp.12421

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*(2), 118–134. https://doi.org/10.1016/j.ijpsycho.2013.08.012

Rosenfeld, J. P., & Labkovsky, E. (2010). New P300-based protocol to detect concealed information: Resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. *Psychophysiology*, *47*(6), 1002–1010. https://doi.org/10.1111/j.1469-8986.2010.01024.x

Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The complex trial protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology*, *45*(6), 906–919. https://doi.org/10.1111/j.1469-8986.2008.00708.x

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*(2), 205–219. https://doi.org/10.1111/j.1469-8986.2004.00158.x

Rosenfeld, J. P., Ward, A., Drapekin, J., Labkovsky, E., & Tullman, S. (2017). Instruction to suppress semantic memory enhances of has no effect on P300 in a concealed information test (CIT). *International Journal of Psychophysiology*, *113*, 29–39. https://doi.org/10.1016/j.ijpsycho.2017.01.001

Stephens, E., Braid, A., & Hertel, P. T. (2013). Suppression-induced reduction in the specificity of autobiographical memories. *Clinical Psychological Science*, *1*(2), 163–169. https://doi.org/10.1177/2167702612467773

Sutton, S., Braren, M., Zubin, J., & John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Verschuere, B., Suchotzki, K., & Debey, E. (2014). Detecting deception through reaction times. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 269–291). John Wiley & Sons, Ltd.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, *26*(2), 208–221. https://doi.org/10.1111/j.1469-8986.1989.tb03159.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Afzali, M. U., Jones, R. D., Seren-Grace, A. P., Palmer, R. W., Makarious, D., Rodrigues, M. N. B., & Neumann, E. (2023). Classification accuracy of the event-related potentials-based *Brain Fingerprinting* and its robustness to direct-suppression and thought-substitution countermeasures. *Applied Cognitive Psychology*, 1–16. https://doi.org/10.1002/acp.4050