

'BRAIN FINGERPRINTING': A Critical Analysis

J. Peter Rosenfeld, Ph.D.

Northwestern University

Recent efforts by various investigators have been directed at using brain waves in detection of deception. One investigator, Lawrence A. Farwell, left academia about a decade ago and founded his company, Farwell Brain Fingerprinting. (See <http://www.brainwavescience.com/>.) This business is actively commercializing a putative deception-revealing technology in connection with forensic and related areas. As is stated on its Web site, "Farwell Brain Fingerprinting is a revolutionary new technology for solving crimes, with a record of 100% accuracy in research with U.S. government agencies and other applications. . . . The technology is fully developed and available for application in the field." The present review undertakes a careful analysis of these claims and their background with reference to the one refereed publication in a major psychophysiology journal which Farwell coauthored (Farwell & Donchin, 1991), the voluminous material on the Web address cited above, Farwell's U.S. patents, court records, and the work of various other researchers. Prior to this analysis, the present review briefly discusses the P300 event-related brain potential, which is the key element of most of the published brain-wave-based deception research. The "Guilty Knowledge Test," or GKT, which in

a form modified for P300 methods yielded the P300 protocol for detecting concealed, crime-related information, is also reviewed, followed by a review of the P300-based deception-detection literature. The issue of P300-based tests' reported accuracies is also considered. Since Farwell claims that his method is based on a brain-activity index, the "MERMER," which goes beyond P300, an attempt is also made to analyze this variable, based, of necessity, solely on U.S. patent material. The review then closely examines Farwell's recently promoted retrospective application of BF, in which the technology is purportedly utilized to exonerate previously convicted—some long ago—felons. Finally, the review highlights methodological problems associated with BF and related methods, including vulnerability to countermeasures and difficulty with developing adequate and appropriate test material, leading to the concluding impression that the claims on the BF Web site are exaggerated and sometimes misleading. It is documented that, in fact, U.S. government agencies most concerned with detecting deception do not envision use of BF. Finally, prospective users and buyers of this technology are issued the classic *caveat emptor*.

J. Peter Rosenfeld, Ph.D., Professor, Northwestern University.

Correspondence concerning this article should be addressed to J. Peter Rosenfeld, Ph.D., Department of Psychology, Institute for Neuroscience, Northwestern University, Evanston, IL 60208.

INTRODUCTION

In the past few years, Lawrence A. Farwell, Ph.D., has been commercializing a putative crime-solving technology closely related to detecting deception that he

calls “Farwell Brain Fingerprinting.” (See <http://www.brainwavescience.com/>.) There has been considerable publicity about this method, most of it cited on the Web page. If one peruses this page, one can see that much of the media have uncritically accepted it with considerable excitement. *Time* magazine has selected Dr. Farwell for the “Time 100: The Next Wave, the Innovators Who May Be the Picassos or the Einsteins of the 21st Century” (although there has also been negative criticism, e.g., http://www.forensicevidence.com/site/Behv_Evid/brainfp_Iowa.html).

As of August 21, 2004, another of Farwell’s Web sites implies that the technique has perfect accuracy by using the phrase “100% accurate” as the subheading of large sections of text (e.g., <http://www.brainwavescience.com/Chemistry.php>). Previous recent versions of his home page claimed this accuracy level very directly: “Farwell Brain Fingerprinting is a revolutionary new technology for investigating crimes and exonerating innocent subjects, with a record of 100% accuracy in research on FBI agents, research with US government agencies, and field applications” (from <http://www.brainwavescience.com/>, 2003). On his main Web site, as of August, 21, 2004, this claim was somewhat qualified: “Farwell Brain Fingerprinting is a revolutionary new technology for solving crimes, with a record of 100% accuracy in research with US government agencies and other applications.” In the present review, analysis will demonstrate that there are many statements on these Web sites that can be questioned, and examples will be provided below. It is appropriate to first review the scientific history and background of the Brain Fingerprinting (BF) enterprise.

BACKGROUND

Farwell claims presently that the brain-wave index crucial to all his assertions is the MERMER, or “Memory and Encoding Related Multifaceted Electroencephalographic Response.” He claims that the P300 event-related potential (ERP, discussed below) is but one element of the MERMER. It will be seen later that P300 is very likely the basis and essence of the MERMER. Indeed, at the Harrington appeal hearing of 2000 (*Harrington v. Iowa*; see http://www.judicial.state.ia.us/supreme/opinions/20030226/010653.asp#_ftnref6), which Farwell claims was the venue in which BF was admitted to court (“Brain Fingerprinting Testing Ruled Admissible in Court;” see <http://www.brainwavescience.com/Ruled>

[20%Admissable.php](http://www.brainwavescience.com/Ruled)), the state’s attorney ultimately forced Farwell to concede that while P300 was a well-established scientific phenomenon—though not necessarily the basis of a successful lie detector—there was no independent, published, peer-reviewed supporting literature on the MERMER. (Elsewhere in Farwell’s many Web pages, the court admissibility issue is handled more cautiously with the caveat: “We believe that Brain Fingerprinting will be admissible in court once the necessary test cases have been tried.” Also: “The admissibility of Brain Fingerprinting in court has not yet been established.”) In any case, it seems unlikely that Farwell would argue against the assertion that the P300 ERP was the brain wave which first impelled several investigators to study the potential of EEG waves as deception indices. The history of this ongoing research program will make this clear. First, however, a brief review of P300 phenomenology is in order.

It is well known that between an electrode placed on the scalp surface directly over brain and another electrode connected to a relatively neutral (electrically) part of the head (i.e., remote from brain cells, such as the ear-lobe), an electrical voltage, varying as a function of time, exists. These voltages comprise the spontaneously ongoing electroencephalogram or EEG, and are commonly known as brain waves. If during the recording of EEG, a discrete stimulus event occurs, such as a light flash or tone pip, the EEG breaks into a series of larger peaks and troughs lasting up to two seconds after the stimulus. These waves, signaling the arrival in cortex of neural activity generated by the stimulus, comprise the wave series called the ERP, the EEG potential series related to the stimulus event.

Actually, the ERP “rides on” the ongoing EEG, by which it is sometimes obscured in single trials. Thus, one typically averages the EEG samples of many repeated presentation trials of either the same stimulus or stimulus category (e.g., male names), and the ensuing averaged stimulus-related activity is revealed as the ERP, while the non-stimulus-related features of the EEG average out, approaching a straight line. P300 is a special ERP that results whenever a *meaningful* piece of information is *rarely* presented as a stimulus among a random series of more frequently presented, non-meaningful stimuli. For example, Figure 1 shows a set of three pairs of superimposed ERP averages from three scalp sites (called Fz, Cz, and Pz) on a single subject, who was viewing a series of test items on a display screen (from Rosenfeld et al., 2004). On 17% of the trials, a meaningful item (e.g., the subject’s birth date)

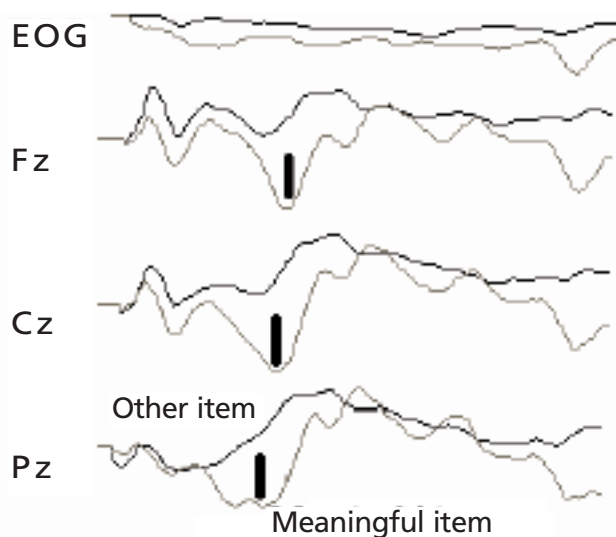


Figure 1. Three ERPs, based on our Rosenfeld et al. (2004) report, from the scalp sites Fz (frontal), Cz (central), and Pz (parietal). The sweeps are 2048 ms. long. P300 peaks are down-going and indicated with thick vertical lines. They are in response to meaningful information items. They are superimposed on responses to meaningless items. Given that the sweeps are about 2 s. long, the P300s begin around 400 ms. and end around 900 ms. Positive down.

was presented, and on the remaining 83% of the randomly occurring trials, other items with no meaning to the subject (e.g., other dates) were presented. The two superimposed waveforms at each scalp site represent averages of ERPs to (1) meaningful items and to (2) other items. In response to the meaningful items, a large down-going P300 (indicated with thick vertical lines) is seen, which is absent in the superimposed other waveforms. (The wave labeled “EOG” is a simultaneous recording of eye-movement artifact activity. As required for sound EEG recording technique, these waves are flat during the segment of time when P300 occurs, indicating that no artifacts due to eye movements are occurring, which, if present, could account for apparent P300s.) Clearly, the *rare, recognized, meaningful* items elicit P300, the other items do not. (Note that electrically positive brain activity is plotted down, as it is traditionally.) It should be evident that the ability of P300 to signal the involuntary recognition of meaningful information suggests that the wave could be used to signal “guilty knowledge” ideally known only to a guilty perpetrator and to police.

EARLY P300-BASED DECEPTION DETECTORS: THE ACCURACY ISSUE

Fabiani, Karis, and Donchin (1983) showed that if a list of words, consisting of rare, previously learned (i.e., *meaningful*) and frequent novel words were presented one at a time to a subject, the familiar, previously learned words but not the others elicited a P300. As suggested above, Rosenfeld, Nasman, Whalen, Cantwell, and Mazzeri (1987) recognized that the Fabiani et al. (1983) study suggested that P300 could be used to detect concealed guilty knowledge, i.e., P300 could be used as a potential lie detector. Therefore, P300 could index recognition of familiar items even if subjects denied recognizing them. From this fact, one could infer deception. The P300 would not represent a lie *per se* but only a recognition of a familiar item of information, the verbal denial of which would then imply deception. Farwell has also emphasized this distinction on his Web site, although as an academic nicety that in no way affects the claims of the BF approach. Farwell and Smith (2001), however, seem to have overextended this distinction: “Brain MERMER testing . . . has almost nothing in common with ‘lie detection’ or polygraphy. Polygraphy is a technique of interrogation and detection of deception. . . . Brain MERMER testing does not require any questions of or answers from the suspect. The subject neither lies nor tells the truth during the procedure, and in fact the results of MERMER testing are exactly the same whether the subject lies or tells the truth at any time.”

This assertion is misleading: in fact, the subject *does* give behavioral button-press responses. One button means “No, I don’t recognize this stimulus.” If the guilty subject presses this no button to a guilty knowledge item, he is lying with his button press, if not his voice. Lying is the clear inference if there is no other innocuous explanation for the brain response, and there is no doubt that P300/MERMER testing is clearly relevant to lie detection. Indeed, the terms “interrogative polygraphy” and “lie detection” are in the subtitle of Farwell and Donchin (1991), Farwell’s only peer-reviewed paper on P300-based deception detection in a psychology, neuroscience, or psychophysiology journal. Finally, when Farwell and Smith (2001—not a journal in psychology, psychophysiology, or neuroscience) stated, “in fact the results of MERMER testing are exactly the same whether the subject lies or tells the truth,” they are incorrect (about the major P300 element of MERMER), and, not surprisingly, did not cite any supportive literature. In fact, there are many peer-reviewed, published studies in which the *opposite* is shown, and it is discussed why

truthful subjects in fact produce much larger P300s than subjects giving dishonest responses to the same questions (e.g., Ellwanger, Rosenfeld, Hankin, & Sweet, 1999; Miller, Rosenfeld, Soskins, & Jhee, 2000; Rosenfeld, Rao, Soskins, & Miller, 2003).

Soon after seeing Fabiani et al. (1983), our lab planned and executed a study (Rosenfeld, Cantwell, Nasman, Wojdak, Ivanov, & Mazzeri, 1988) in which subjects pretended to steal one of ten items from a box. Later, the items were repeatedly presented to the subject by name, one at a time, on a display screen, and we found that the items the subjects pretended to steal (the *probes*), but not the other, *irrelevant* items, evoked P300 in 9 of 10 cases. In that study, there was also one special, unpredictably presented stimulus item, the *target*, to which the subjects were required to respond by saying "yes," so as to assure us they were paying attention to the screen at all times and would thus not miss probe presentations. They said "no" to all the other items, signaling nonrecognition, and thus lying on trials containing the pretended stolen items. The special target items also evoked P300, as one might expect, since they too were rare and meaningful (task-relevant). (The 1988 study was actually the second of two closely related publications, the first having been published as Rosenfeld et al., 1987.) This paradigm had many features of the guilty-knowledge test (GKT) paradigm (developed by Lykken in 1959; see Lykken, 1998), except that P300s rather than autonomic variables were used as the indices of recognition. This required various other departures from the classic GKT method, such as signal averaging and target stimuli.

Donchin and Farwell also saw the potential for detecting deception with P300 as a recognition index in the later 1980s, and they presented a preliminary report of their work (in poster format) at the 1986 Society for Psychophysiological Research (SPR) meetings. This conference abstract summarized experiment 2 of the paper later published as Farwell and Donchin, 1991, which Farwell cites as a solid empirical foundation of the BF method. Indeed, it is, as already noted, the sole peer-reviewed publication involving Farwell in a leading psychophysiology/psychology/neuroscience journal supporting detection of concealed information with P300—although not with MERMER (see below). Although there is the aforementioned latter publication by Farwell and Smith (2001), cited on the BF Web sites, as was already noted, this later paper appeared in an outlet which is not a peer-reviewed or leading journal in psychology, neuroscience, or psychophysiology. Indeed, it is unlikely that this report would have appeared in a

major journal, since the detailed stimulus lists, MERMER methods, and results were undisclosed, and no major scientific journal will accept such a paper since it is impossible to replicate with key methods kept secret. Also curious about the late negative component said to be a part of MERMER (see below), Farwell and Smith (2001) stated that it is largest over the frontal cortex (site Fz), yet they show no frontal waveforms, even though these frontal waveforms were said to have been recorded. Moreover, only three guilty and three innocent subjects were run. This is an unacceptably low number of subjects for a high-quality, peer-reviewed publication. Indeed, the authors acknowledge this significant limitation: "It would be inappropriate to generalize the results of the present research because of the small sample of subjects," but this qualification appeared only in the discussion section of the paper on the Web site. Although Farwell and Smith (2001) claimed that 100% of the subjects were accurately identified, actually five were identified with 99% confidence, and one with only 90%. (In psychology, one usually likes to see at least 95% confidence levels.) One also learns from a careful study of this report that of these six subjects, one replaced a discarded subject because of "the [discarded] subject's not understanding, and consequently, not following the instructions." It would seem that real test subjects in the field could also have such problems, so that generalizing from this paper to real field application becomes problematic. Let us closely examine, therefore, the original study of Farwell and Donchin (1991), also a guilty-knowledge paradigm (Lykken, 1998), which utilized P300s as the physiological variables indexing recognition of items of concealed guilty knowledge.

This study reported on two experiments, the first of which was a full-length study using 20 paid volunteers. The second experiment contained only four subjects, and we shall look at it momentarily. In both experiments, subjects saw three kinds of stimuli, quite comparable to our Rosenfeld et al. (1988) study, noted above. There were *probe* stimuli that were items of guilty knowledge, which only "perpetrators" and authorities (experimenters) would have. Then, there were *irrelevant* items that did not relate to the "crime." Finally, there were target items, as above, which were irrelevant items, but to which the subject was instructed to execute a unique response. That is, the subjects were instructed to press a yes button for the targets, and a no button to all other stimuli. The subjects participated in a mock-crime espionage scenario, in which briefcases were passed to confederates in operations that had particular names. The details of these activities generated six categories of

stimuli, one example of which would be the name of the mock espionage operation. For each such category, the actual probe operation name might be operation “donkey.” Various other animal names—tiger, cow, etc.—would comprise the rest of the set of six stimuli, including the probe, four irrelevants and one target name. The six (categories) times six (one probe+one target+four irrelevant stimuli)=36 items, which were randomly shuffled and presented twice per block. After each block, the stimuli were reshuffled into a new random order and re-presented for a total of four blocks. The mock crime was committed one day before the P300 GKT. *Very important: prior to the P300 GKT, indeed, prior to performance of the mock-crime scenario, each subject was tested and trained on the details of the mock crime in which he/she participated. The training was to a 100% correct criterion. Thus, the experimenters could be quite certain that the crime details would be remembered.* Subjects were also trained to know the targets. Subjects were also run as their own innocent controls by having been tested on scenarios of which they had no knowledge.

Farwell and Donchin (1991) reported that in the 20 guilty cases, correct decisions were possible in all but two cases, a detection rate of 90%. Indeed, this was not impressive, given that the subjects were *trained* to remember the details of their crimes, *a procedure having limited ecological validity in field circumstances*—in which training of a suspect on details of a crime he/she was denying would not be possible. In the innocent condition, only 85% were correctly classified, yielding an overall detection rate of 87.5%.

In the second experiment of Farwell and Donchin (1991), the four volunteering subjects were all previously admitted wrongdoers on the college campus. Their crime details were well-detected with P300, *but these previously admitted wrongdoers no doubt had had much review of their crimes at the hands of campus investigators, teachers, parents, etc. Therefore, one can ask: was the P300 test detecting incidentally acquired information—versus previously admitted, well-rehearsed information?* Moreover, the $n=4$ was hardly convincing, and in one of the four innocent tests, no decision could be rendered, meaning that a correct decision was possible in only three of four (75%) innocent cases.

How does this sole peer-reviewed empirical foundation square with BF’s implied claims of 100% accuracy or with the following statement on Farwell’s Web site? “Farwell Brain Fingerprinting is based on the principle that the brain is central to all human acts. In a criminal act, there may or may not be many kinds of peripheral

evidence, but the brain is always there, planning, executing, and recording the crime. The fundamental difference between a perpetrator and a falsely accused, innocent person is that the perpetrator, having committed the crime, has the details of the crime stored in his brain, and the innocent suspect does not.”

This assertion contains two implications that are clearly open to challenge: (1) Perpetrators are always planning their crimes. This contention is easily disputed; in fact, many serious crimes are unplanned and impulsive. (2) The brain is constantly storing undistorted, detailed representation of experience that the BF method can extract from the brain just as easily as real fingerprints can be lifted from murder weapons (hence the misleading term, “Brain Fingerprinting”). Regarding the critical second implication, it is well known from the memory literature that, in fact, not all details of experience are recorded, or if recorded, then often recorded with major distortion; the fragility of memory is well documented (e.g. Ford, 1996, pp.174–176; Loftus & Loftus, 1980; Loftus & Ketcham, 1994). Moreover, it is likely that an individual in the act of committing a serious crime—from murder to bank robbery to terror bombing—would be in such an excited or anxious state so as to render his/her attention to details of the crime scene close to inoperable. Also, a high proportion of crime in the U.S. is committed under the influence of drugs or alcohol, which are known to play havoc with memory. Finally, the P300-based detection of well-rehearsed incidental knowledge items—the types of details from which BF’s probe stimuli would be composed—is rather poor in comparison to detection of high-impact, over-rehearsed, autobiographical knowledge (Rosenfeld, Biroshack, & Furedy, 2005). Indeed, using the three-stimulus protocol discussed above, only 40% of the subjects were detected and identified as having acquired incidental knowledge.

It must again be borne in mind regarding the immediately preceding quotation that a criminal suspect in the field would hardly have had the kind of rehearsal opportunities present in both experiments of Farwell and Donchin (1991). Indeed, if well-rehearsed subjects are detected 90% of the time (18 of 20 in Farwell & Donchin, 1991, experiment 1), that does not support the notion that “the brain is always there, planning, executing, and recording the crime.” But it is well-known that the brain seems *not* to be always there, or if it is always there, it is *not always attending* to the BF-appropriate crime details. *Very important:* the empirical foundation just described does not support the current *retrospective* use of BF, a use that is strongly encouraged by the BF

Web site, in which long-incarcerated convicts are tested with BF on stimuli developed decades after the crime. This questionable application is analyzed below.

Regarding P300-based GKT studies from independent laboratories, how does the BF method fare? Our lab has typically reported 80–95% detection. (See Rosenfeld, Soskins, Bosh, & Ryan, 2004; Rosenfeld, 2002.) Our higher detection rates tend to accompany detection of autobiographical knowledge in head-injury-malingering studies. Our lower rates tend to accompany detection of incidental knowledge as in Rosenfeld et al., 2004. Another prominent group of investigators used to report 90% and above hit rates in detecting concealed though *over-learned* information, (Allen, Icano, & Danielson, 1992). More recently, this group has reported poor detection (27–47%) of mock-crime details (Mertens, Allen, Culp, & Crawford, 2003). Apart from these lab analogues, there has been only one independent field study of P300-based detection of guilty knowledge, that by Miyake, Mizutanti, and Yamahura (1993). This study, under the auspices of a Japanese police department, reported only 48% detection of guilty subjects.

One can surmise what Farwell's responses to these challenging data would be, based on the fact that he was actually confronted with the Miyake et al. (1993) report at the Harrington 2000 hearing. He stated that these findings were not relevant, since Miyake et al. recorded from Cz rather than Pz: "They recorded from Cz, so I don't know what they were measuring . . . it appears they were doing something that was in no way related to what we did." This statement seems erroneous and misleading, in that Miyake et al. were indeed conducting related research, as they actually cited Farwell and Donchin (1991) as the basis of their effort. Moreover, had there been a P300 expert present, he/she could have retorted that P300s from Cz and Pz usually correlate at $>.95$ over trials, and that, indeed, no less a P300 expert than Polich (1999) recommended the use of Cz in diagnostic clinical P300 studies.

Farwell might also respond more technically that the EEG filters used by other investigators are not the *Optimal Digital Filters* he used in Farwell and Donchin (1991), and claimed to be superior to the filters most others use (Farwell, Martineri, Bashore, Rapp, & Goddard, 1993.) The filters discussed here are circuit elements—or software models of circuit elements—through which raw EEG signals are passed. Their purpose is to remove artifactual and other sources of noise in the brain-wave signal. The present author, not an electrical engineer, had always sensed a problem with the Farwell et al. (1993) paper. In preparing the present review, I consulted two P300 experts (one an engineer),

plus one of Farwell's coauthors on the 1993 paper about this serious problem. Here is what I wrote, quoted directly from my e-mail:

One thing has always disturbed me about that paper by Farwell, Martineri, Bashore and Paul Rapp and Goddard (Psychophys v30, no.3, pp306, 1993). What they did was to take some raw EEG with P300 from 2 Ss and apply boxcars [a standard filter] and then Optimals and showed that the resulting P300s from the Optimals looked bigger than those from the boxcars. Larry Farwell then generalized . . . that here is clear evidence of the superiority of the Optimals. I never figured out how he knows what the "real P300" in the subject's head looked like, after all, he was looking at a mixture of signal and noise [i.e., a real EEG record] in the first place. If I were going to try to conclude as Larry does, I guess I would *start with simulated P300s, add some simulated EEG-like noise* and then test my creation with the two kinds of filters. If the Optimal's output looked more like the *known input signal*, then ok. Otherwise, I am puzzled. Comments? [Bracketed phrases and italics were added in this paper for clarity.]

In other words, if you take an unknown mixture of signal (P300) and noise (background EEG) and apply two different filters and compare their outputs, how can you know which more closely represents the true signal, since the true signal was combined with noise in the first place?

Here is what the experts responded, quoted directly, with their gracious permission, from their e-mails in response to my question:

(1) I agree with you and, for good measure, note that the Optimals required more time points than the boxcars to get comparable bandwidths. This can be a problem if you have limited data epochs—less of the epoch can be filtered. [This was from Daniel Ruchkin, D. Eng., an electrical engineer, author of a classic text on EEG/ERP signal analysis and a highly regarded cognitive neuroscientist.]

(2) I agree with you. You need to do a thoroughly detailed simulation in which you also vary the conditions of the simulation mimicking experimental manipulations and then see which filtering technique works best. [This was from Emanuel Donchin, Ph.D., arguably the leading P300 researcher since the discovery of this ERP and, interestingly enough, the Ph.D. mentor of Lawrence A. Farwell, and thus, coauthor of Farwell and Donchin, (1991).]

Finally, I contacted Paul Rapp, Ph.D., a mathematician at Drexel University, who was one of the coauthors of the Farwell et al. (1993) paper, with the same question. Here is his answer: "Peter, Your assessment is correct and your concerns are fully warranted."

There is therefore no reason to assume that the lack of use of optimal digital filters renders the negative results described by Mertens, Allen and colleagues, Miyake and colleagues, and Rosenfeld and colleagues invalid. The key piece of evidence Farwell needs in order to establish the superiority of these filters *in deception-detection situations* is an experiment in which both boxcars and optimals are applied to the same deception-model data with the results being that a significantly greater proportion of correct decisions results from the use of the optimals. This result would need to be replicated by at least one independent group of researchers before the scientific community would accept it. To the best of the author's knowledge, no such study has been conducted. Until such work is completed, the negative results from other laboratories constitute a lack of support for BF.

Again, one can anticipate a rejoinder from the BF group: "Whether or not our filters have been soundly established in the literature is beside the point. The fact is, we used them and we get our spectacular results, and all the others do not." To this, one can reply only that (1) one usually requires independent validation of a proclaimed superior, novel technique prior to general, scientific acceptance and (2) there are always many small differences among laboratories that could be used to claim that *perfect* replication is never possible. Indeed, Farwell probably uses different biological amplifiers than others do. No doubt, the physical environments of the various laboratories must differ in various ways. If any of these differences is responsible for replication difficulties, *then the best one can say is that the BF paradigm does not seem to generalize well out of Farwell's hands*. Indeed, despite differences among laboratories, dozens of investigators have reliably evoked (non-forensic) P300s in other known ways. Finally, Farwell might attempt to point out that whereas the rest of us use a basic P300 as our brain-wave index of concealed information, he uses what he claims to be the more comprehensive "MERMER" response. Just what is this?

MERMER

Find the MERMER
and you have found
the murderer.

[From the BF Web pages]

What is a MERMER? This is not easy to pin down from Farwell's Web pages, or from elsewhere. In one place, he says, "If the computer detects a brain P300/MERMER, this indicates that specific information relevant to the situation under investigation is stored in the brain" (<http://www.brainwavescience.com/FBIStudy>.

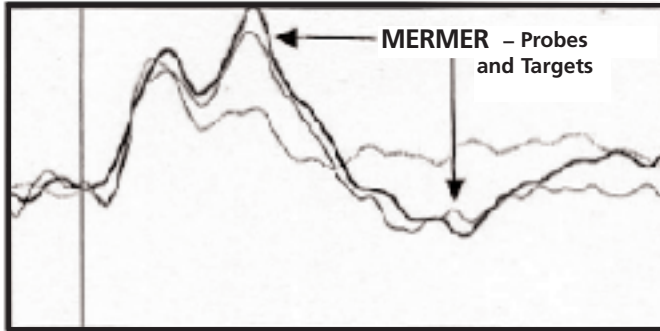
php). This "P300/MERMER" terminology suggests that MERMER is a kind of P300. However, elsewhere on the same page, Farwell, speaking of the Farwell and Donchin (1991) data, states: "87.5 percent of the subjects were correctly classified as having or not having the relevant information. The remaining 12.5 percent were indeterminate. Rosenfeld and his colleagues (10, 11, 12, 13), and Allen and Iacono (14, 15) achieved comparable results with similar procedures using ERPs. Present-day brain MERMER testing, including the data reported here as well as in four other studies by Farwell and his colleagues (9, 16, 17, 18, 19, 20, 21), has achieved an even higher level of accuracy than that achieved in the ERP studies. In all five of these MERMER studies, accuracy has been 100 percent with no false negatives, no false positives, and no indeterminates." (This quotation is also in Farwell and Smith, 2001.)

This quotation states directly that by using P300 alone, one can correctly classify only 85–90% of the cases, but with MERMER, one can approach perfection. *The problem is that none of these seven references cited to support MERMER's high accuracy are published, peer-reviewed studies in any journal*. They include three patents, two private presentations to the CIA, a meetings abstract, and Farwell's unpublished doctoral thesis. They appear to cover the same overlapping data sets. In any case, of what does MERMER actually consist? Here's Farwell's answer from the BF Web pages:

One of the most easily measured aspects of this [MERMER] response (and the only one measured in early research) is an electrically positive component, maximal at the midline parietal area of the head, with a peak latency of approximately 300 to 800 ms. It is referred to variously as P300, P3, P3b, or late positive component (LPC). Another more recently discovered aspect of the MERMER is an electrically negative component, maximal at the midline frontal area, with an onset latency of approximately 800 to 1200 ms. These components can be readily recognized through signal averaging procedures. Recent research suggests that a third aspect of the P300/MERMER is a pattern of changes in the frequency domain characterized by a phasic shift in the frequency power spectrum that can be detected using single-trial analysis techniques."

Again, one wonders: if MERMER is more than P300, why again write, "P300/MERMER?" Also, quite parenthetically, when Farwell lumps P300, P3b, and the late positive component, or LPC, together as if they are all the same thing, he ignores or is unaware of the rather considerable controversy which has existed for many years about this putative equivalence. See for example, Spencer, Dien, and Donchin, 2001).

A. Parietal Area



B. Frontal Area

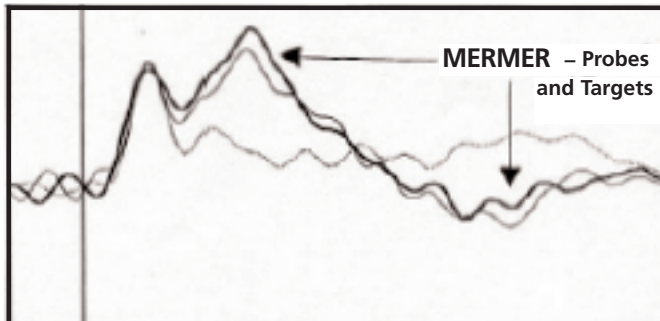


Figure 2. These are ERPs that are copied from the BF Web site, showing Pz (top) and Fz (bottom) ERPs. Positive is up, and the "MERMER" in response to targets and probes indicates two peaks with arrows; the first is the classic P300, and the second is the subsequent negative peak. The unlabelled waveforms not containing "MERMERS" are in response to irrelevant stimuli. (No timing information provided.)

Figure 2A, copied directly from Farwell's Web pages, illustrates the MERMER as recorded from Pz. One notes that arrows point to both an up-going wave and a down-going wave that follows. The first wave is clearly the classic P300. (Note that positivity is plotted upwards here.) The second wave is that "more recently discovered aspect of the MERMER . . . an electrically negative component . . . with an onset latency of approximately 800 to 1200 ms." In fact, our lab has been utilizing this "more recently discovered" negative wave, which we finally, formally dubbed NEG (in Soskins, Rosenfeld, & Niendam, 2001) in our P300-based deception studies, at least since our 1991 study of P300-based detection of concealed knowledge (Rosenfeld, Angell, Johnson, & Quian, 1991). We argued for many years prior to the appearance of MERMER that the purer and usual measure of P300 taken from prestimulus baseline, the standard baseline-to-peak or b-p measure, is consistently *less accurate in deception studies* than the peak-to-

peak (p-p) index of P300 measured as the difference from NEG. Thus, our often-used p-p index of P300 combines P300 with the subsequent negative component, perhaps not appreciably differently than MERMER does. Perhaps MERMER combines both P300 and NEG with some regression-based equation more elaborate than our simple linear difference, but since Farwell's specific method of combining P300 and NEG into MERMER is undisclosed (even in his patents), unpublished, un-replicated independently, and not peer-reviewed, one cannot be persuaded that this is the new and putatively independent feature of MERMER that makes it near perfect: the p-p index combining both P300 and NEG has been good (80–95% correct classification)—but not perfect. Indeed, there is a possible reason for this: Soskins et al. (2001) conducted a systematic study of NEG and confirmed that it partly represented a capacitive rebound artifact of the filter settings used, and partly represented the recovery time to baseline from the positive P300 peak. This latter variable is potentially independent of P300 amplitude, and thus may represent novel attributes of the P300 process. This could account for the fact that the p-p measures are more accurate than the b-p measures in detection of deception, but even with it, we cannot achieve >95% detection, and usually report 80–88% in lab analogues of crime situations.

Another concern: in one of his more recent patents (Farwell, 1995), Farwell emphasized that what we call NEG is a mostly *frontal* (Fz) component. If this claim is accurate, then Farwell's negativity differs from our NEG, which is related to P300, as described above, and, in our hands, is *parietally maximal along with P300 to which it is related*. However, this claimed *Fz dominance has never been clearly shown by Farwell (nor independently replicated)*. In Farwell and Smith (2001), only Pz responses are shown, and the negative wave is quite apparent in this publication at Pz. Even though Farwell reported recording from Fz in that paper, Fz data are not presented. In Figure 2 (from the BF Web site—<http://www.brainwavescience.com/FBIStudy.php>) above, the Pz response is in 2A and the Fz response is in 2B. The negative wave for probes and targets, relative to irrelevant, appears slightly though not necessarily significantly *larger* at Pz than at Fz. *Certainly, there is no Fz dominance*. In our data (Figure 1 is but one of dozens of possible examples), our NEG wave at Fz is usually smaller than at Pz (see Rosenfeld et al., 2004). If the *frontal* negativity stated by Farwell to be a critical feature of MERMER, independent of the Pz-maximal P300, then one wonders why it is so difficult to find an example of this claimed Fz dominance in all of his writings, including patents, and why, using

very similar paradigms, don't others see this putative Fz dominance? Shouldn't this claimed novel and critical MERMER element be showcased?

What about the third aspect of MERMER? Farwell stated that: "Recent research suggests that a third aspect of the P300/MERMER is a pattern of changes in the frequency domain characterized by a phasic shift in the frequency power spectrum that can be detected using single-trial analysis techniques." (To the best of the author's knowledge, the "recent research" mentioned is not published.) What he is saying, in part, very simply, is that a P300-containing trial has more lower-frequency activity than a trial lacking P300. (For example, in his 1995 patent [Farwell, 1995], he speaks of ". . . the very slow activity in the range of 0.1 to 2 Hz that contributes to the MERMER.") But this is self-evident: P300 is a fairly low-frequency ERP with a duration of about .5 s., meaning a mean frequency of 2 Hz. EEG (from an alert subject) not containing P300 contains frequencies mostly between 5 and 50 Hz. (depending on the subject's alertness). It follows that this analysis in the frequency domain that comprises the third element of MERMER is very likely simply yet another way to detect P300.

About these time-domain elements of MERMER, Farwell's (1995) patent further states a few paragraphs later: "The frequency domain characteristics of the MERMER comprise *at least one of the following*: an increase in power from .1 to 4 Hz., a decrease in power from 8 to 12 Hz., and an increase in power from 12 to 20 Hz.. It may be appreciated that *the pattern may differ for different subjects. . .*" [Present author's emphasis.] From this, one gathers that *the MERMER frequency characteristics are not the same in all subjects*. We have come a long way from the basic P300, which is usually similar across subjects. There is good theory to explain what the P300 is and why one would expect it in most subjects recognizing rare, meaningful probes. The meaning of these other claimed independent (but undocumented) frequency phenomena, which, according to Farwell himself, are not found in all persons, is another matter. Farwell might argue (and appears to do so in the patent) that it does not matter if one has a different set of spectral changes from one subject to the next; subjects are always their own controls, as within the subject, one examines whether or not the probe-irrelevant MERMER is larger than the probe-target MERMER. But long-established theory tells us what to expect in the case of P300 alone: if the probe P300 looks more like the target P300—the latter of which is the built-in, exemplar control for what one expects to see in a subject recognizing

meaningful information—then the probe-evoked P300 response also represents a response to meaningful information which can happen only if the subject can discriminate the probe. *About the putative spectral changes, we know nothing*. Thus, if the probe-target spectral change correlation in a subject is greater than the probe-irrelevant spectral change, one has no idea what this means. Indeed, the statement above about the three types of spectral changes one may sometimes see in a MERMER, quoted from the 1995 patent, is a simple summary statement. *The supportive data—e.g., power spectra illustrating these claimed frequency effects—have never been shown anywhere*.

This leads to the important observation that nowhere from Farwell is there any explanation of what the new MERMER components—the negative component and the frequency analysis—signify psychologically or physiologically. Indeed, even if Farwell does not want to disclose, for proprietary reasons, the function which includes P300, NEG, and the frequency analysis, it would certainly be useful for him to inform the scientific community of what proportion of the variance of MERMER is orthogonally accounted for by each of its three elements. That is, how much, quantitatively, do the frequency analysis and the late negative component (NEG) each add to the MERMER's accuracy? This remains an unanswered question.

Again, there is no published or independently replicated information detailing how P300, NEG, and frequency analysis are combined into MERMER. If there are no published details about how to generate a MERMER, then there can never be independent replication.

RETROSPECTIVE APPLICATION OF BF: THE HARRINGTON CASE

The BF Web page cites the Harrington case (*Harrington v. Iowa*, 2000) as the one that was admitted to court and that illustrates how BF can be used to exonerate wrongly convicted, incarcerated persons. By way of background, one should be aware that Harrington was sentenced to life in prison over 20 years ago for the murder of a security guard. Two decades later, his attorneys appealed for a reversal on several grounds, mostly including suppression of evidence by authorities and recantation by dubious witnesses, but also including BF evidence. (See *Harrington v. Iowa*, 2000; see also http://www.judicial.state.ia.us/supreme/opinions/2003/0226/01-0653.asp#_ftnref6. This is the final ruling by the Iowa Supreme Court. It contains references to the

earlier proceedings.) In his contribution to the recent appeal, Farwell constructed a set of probes, targets, and irrelevant details based on the facts of the crime, including incidental details of the murder scene, the getaway route, and so forth. When tested on these, Harrington apparently had no memory of the crime details, leading to the naïve inference that being innocent, he wasn't there and couldn't store the scenario details. This inference was naïve because notwithstanding the fragility of incidental memory retention, particularly during a highly emotionally charged act, as discussed above, *the stimulus set was developed and the test was administered more than 20 years after the event!*

Moreover, the details of the ERP analysis—specifically, the epoch of waveforms over which the analysis was performed—were not disclosed. Actually, when the ERPs based on the guilty scenario (as submitted by Farwell to *Harrington v. Iowa*, 2000) were first published on the BF Web site, the P300s from the Harrington tests revealed *two* positive peaks in the region where P300 normally appears, that is, between 300 and 800 ms. post-stimulus—the P300 range according to the quotation above from the BF Web site, and with which most P300 experts would agree. (See Figure 3, from the report that Farwell submitted to the 2000 Harrington appeal hearing.) The first peaks of probes and targets were of about the same size. The second peak of the target (labeled P300-2 in Figure 3 of the present paper) was distinct, but it was absent in the probe waveform. Thus, if one analyzed the region containing both peaks (as in the author's opinion, one should have done), the results could have been incriminating. If one's analysis started between the peaks, then the "information absent" (innocent) decision would result (as shown in Figure 4, left panel). Since it is nowhere detailed, one does not know what Farwell's formal analysis did in this regard. It is likely that any disinterested P300 researcher would be hard-pressed to make

the arbitrary decision that in Figure 3, one peak was and the other wasn't P300, since both occurred well within the P300 latency range (see the time scale at the bottom of the figure), and since such multiply peaked P300 averages are quite common. It is also the case that later pictures of the waveforms on the BF Web sites do not start at (or actually 100–200 ms. before) the stimulus marking the beginning of the epoch as is customary in peer-reviewed papers (so as to show the pre-stimulus baseline EEG level), but indeed commence *between* the waves noted above (Figure 4). Indeed the P300 averages on the top left and top right of BF's home page clearly do not begin at stimulus presentation, since they are offset from baseline. The present detailed description one finds on the BF Web page, at <http://www.brainwavescience.com/Chemistry.php>, includes the figures shown here as Figure 4, left, titled by Farwell, "Harrington's get out of jail card," i.e., the results of the test of the guilty-crime scenario. (This material was printed from the Web page on August 31, 2004, then scanned into a file which was changed to grayscale, relabeled according to the BF Web page legend, and then

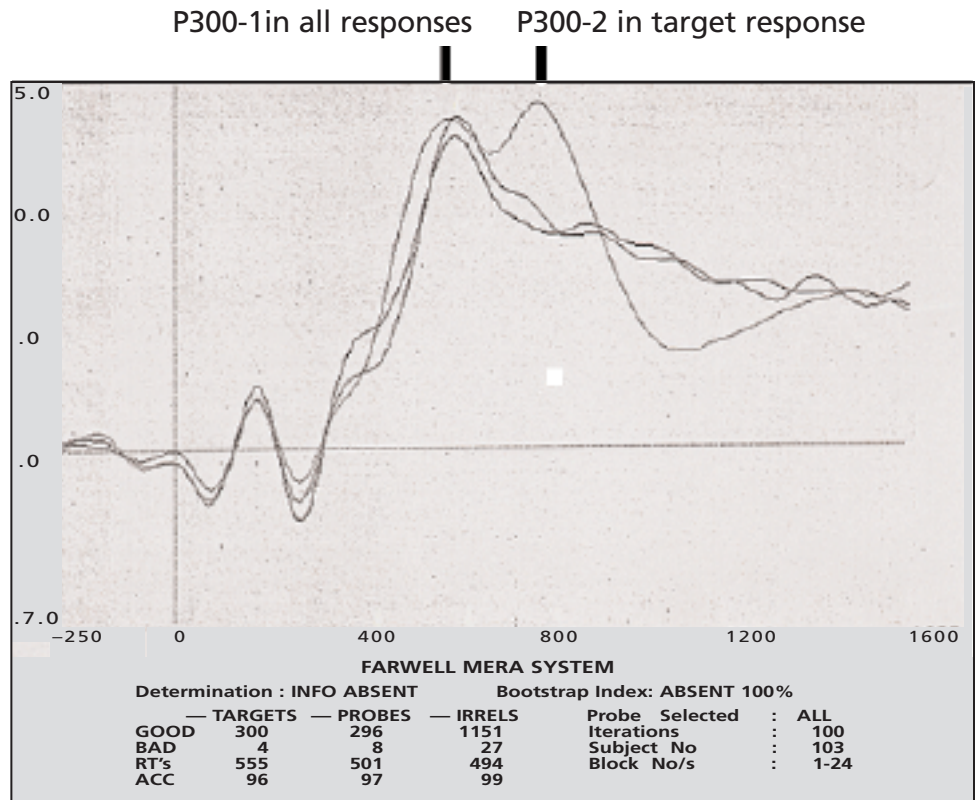


Figure 3. This is from the report of the BF test on Harrington (*Harrington v. Iowa*, 2000). The probes were based on the crime scenario. I have indicated an early P300 (P300-1) seen in all responses and a late P300 (P300-2) seen only in target responses. The early P300 begins around 400 ms. (like ours in Figure 1) and ends at 600ms., where the second peak begins. (If one closely inspects the Pz response from our Figure 1, one can observe a doubled peak also, although they are not as separated as in this figure.) Positive is up.

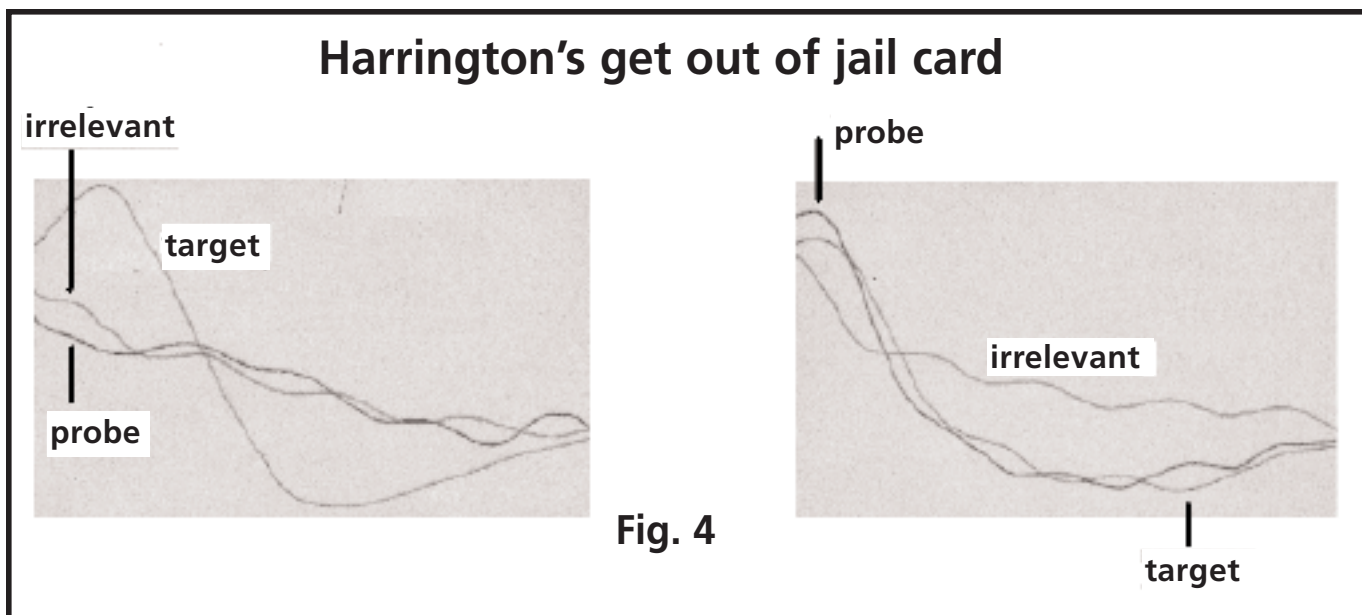


Figure 4. This is what is presently on the BF Web page as indicating an information-absent (in the crime scenario) diagnosis at left (the entire waveform set is in Figure 3) and an information-present diagnosis in the alibi scenario at right (the entire waveform set is in Figure 5). Positive is up.

pasted here as Figure 4, left.) It is obvious that only the latter part of the uncut, actual, entire waveform in Figure 3 is presented in Figure 4, left panel. This part would seem very exculpatory as BF claims, since it appears that only the target response contains the P300 and subsequent negative wave (NEG). However, it does not include the earlier peak—which, most experts would agree, is also part of P300.

Farwell also constructed a BF test and tested Harrington (again more than 20 years after the fact) on his “alibi scenario.” That is, Harrington testified that he had been to a restaurant and concert on the night of the murder, and so the probes, targets, and irrelevants were taken from this scenario. Harrington is said by the BF Web pages to have scored well on that test, in that he had distinct P300s to the probes from that scenario. A naïve interpretation would be that he must have been at that concert and restaurant and is therefore innocent of the crime. An enlightened interpretation is that whether guilty or innocent, Harrington probably rehearsed the details of his alibi repeatedly over the many years after his first trial, during his incarceration, and during the appeal process. Thus, again, did the P300s of the alibi scenario signify *recalled* versus *rehearsed* alibi details, both of which would evoke P300? It seems evident that this retrospective use of the brain-wave-based GKT is open to serious challenge on this basis alone.

Yet there are other serious concerns. Figure 4, right panel, shows what’s currently on the BF Web page as the

result of the test on the innocent scenario. It suggests that the probe and target in the alibi scenario evoke a MERMER, but not the irrelevant. Here, it is mainly the NEG wave that shows similarities between probe and target, both of which appear to differ from the irrelevant. However, this picture is truncated also, as we will shortly see—note how it starts well off baseline—but even this truncated picture indicates not much difference in the positive waves—the actual second P300 peak—shown at the beginning of the epoch among the three waveforms. Now, let us look at Figure 5, the entire (i.e., *uncut*) waveform set from the alibi scenario as submitted by the defense to Harrington’s 2000 hearing. There does not appear to be much difference among the three waves. Certainly, the differences in the P300 latency range are about the same as the differences among the waves in the earlier negative components. Indeed, regarding the first positive peak in Figure 5—chopped off on the present BF Web page—it is the *irrelevant* waveform which has the largest positive peak, and *the target has the smallest*, making any “information-present” interpretation (as on the BF Web site seen in Figure 3) based on these basically similar P300 sizes, highly questionable. If the analysis were based solely on NEG, then the subject would seem to know the alibi information—for whatever reason. (See above). But as is reported in Soskins et al. (2001), NEG by itself, just like P300 by itself (i.e., the b-p P300), is a much less reliable index of possessed knowledge than the p-p P300.

In any case, neither Farwell and Donchin (1991) nor Farwell and Smith (2001) provide support for this retrospective use of BF as in the Harrington case, since these studies tested for recently acquired information by well-practiced subjects. There is, in fact, no published, peer-reviewed, scientific evidence whatsoever supporting this retrospective testing as was done in the Harrington case. Moreover, Farwell's selected sections of the key waveforms appearing in his Web pages raise questions about the data presentation and analysis.

It happens that the Iowa Supreme Court did recently grant Harrington a new trial, thus reversing the decision of the appeals court in 2000, which, BF notwithstanding, had denied Harrington's appeal. The BF Web site clearly suggests that the BF evidence was of major importance in that Supreme Court decision. (Recall also the "get-out-of-jail card," noted above and in Figure 4):

Iowa Supreme Court overturns the 24 year old conviction of Terry Harrington, Brain Fingerprinting Test aids in the appeals; Iowa Supreme Court Reverses Harrington Murder Conviction after 24 Years Brain Fingerprinting Test Supports Innocence

In fact, the Iowa Supreme Court decision (http://www.judicial.state.ia.us/supreme/opinions/20030226/01-0653.asp#_ftnref6) makes it clear that the reasons for their overturning the lower court and giving Harrington a new trial had to do with suppressed evidence and dubious, recanted testimony, not BF:

We also think the reports were "suppressed" within the meaning of the Brady rule. . . . We conclude Harrington did not have the "essential facts" of the police reports so as to allow the defense to wholly take advantage of this evidence. . . . Upon our de novo review of the record and consideration of the totality of the circumstances, our collective confidence in the soundness of the defendant's conviction is significantly weakened. Hughes, the primary witness against Harrington, was by all accounts a liar and a perjurer. . . .

Under the circumstances presented by the record before us, we cannot be confident that the result of Harrington's murder trial would have been the same had the exculpatory information been made available to him. We hold, therefore, that Harrington's due process right to a fair trial was violated by the State's failure to produce the police reports documenting their

investigation of an alternative suspect. . . . Accordingly, we reverse the trial court's contrary ruling, and remand this matter for entry of an order vacating Harrington's conviction and granting him a new trial." [See the present Appendix 1 or the Web site given above for more complete information.]

What role did BF actually play in the Iowa Supreme Court decision? It merited one line in this decision, and was then disregarded. (Note especially the last sentence in the following excerpt):

The same police reports, in addition to recantation testimony and novel computer-based brain testing, also served as a basis for Harrington's claim of newly discovered evidence under section 822.2(4). . . . Because

Fig. 4

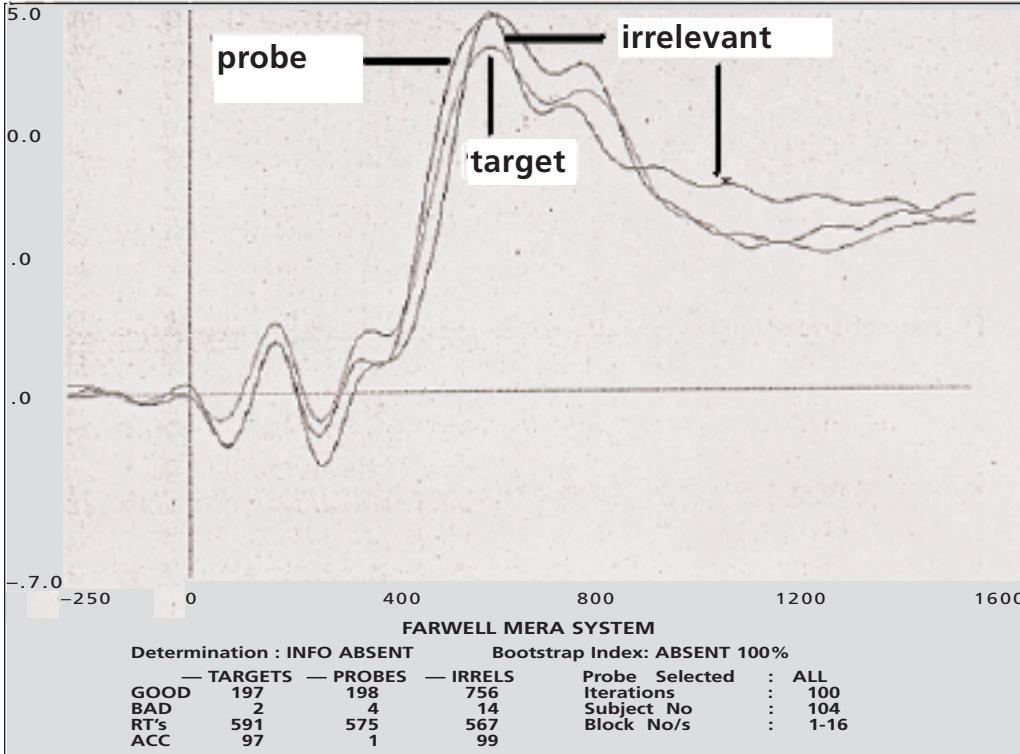


Figure 5. This is the same as Figure 3, except it is the BF report of Harrington's test on the alibi scenario. The most prominent positive peak is from 400–600 ms. It is the main P300, and it is largest in the irrelevant response and smallest in the target. This is not really interpretable. The late negative wave only seems more prominent in probes and targets than in irrelevants. Positive is up.

we conclude the due process claim is dispositive of the present appeal, we do not reach the question of whether the trial court erred in rejecting Harrington's request for a new trial on the basis of newly discovered evidence [BF]. Nonetheless, we briefly review the evidence introduced by the defendant at the PCR hearing with respect to various witnesses' recantation of their incriminating trial testimony, as it gives context to our later discussion of the materiality of the police reports. Because the scientific testing evidence [i.e., BF] is not necessary to a resolution of this appeal, we give it no further consideration. [Bracketed material added by the author.]

COUNTERMEASURES, METHODOLOGICAL, AND ANALYTIC ISSUES

One of the most serious potential problems with all deception-related paradigms based on P300 as a recognition index is the potential vulnerability of these protocols to countermeasures (CMs). These are covert actions taken by subjects so as to prevent detection by a GKT. (See Honts & Amato, 2002; Honts, Devitt, Winbush, & Kircher, 1996.) One might think that CM use would be detectable, and thus not so threatening to P300-based deception detection. For example, if the subject simply failed to attend to the stimuli, then there would be no P300s to the targets, and that would indicate noncooperation. If, as an alternative strategy, the subject started giving special responses to all irrelevant stimuli—as has been reported in ANS-based deception detection (see Honts et al., 1996)—one might expect a systematic change in reaction time (RT), as well as a change in the pattern of P300s evoked in CM users. We recently conducted a formal study of the countermeasure question (Rosenfeld et al., 2004), in which these matters are fully discussed and in which we challenged the Farwell and Donchin (1991) method, as well as our own somewhat different approach (e.g., Rosenfeld et al., 1988, 1991) with CMs. We chose to challenge Farwell and Donchin (1991) rather than Farwell and Smith (2001), with CMs, because, really, there was no choice. First, the earlier report was a solid scientific study in a serious journal and coauthored by Donchin, a highly respected scientist. Second, as already noted, the Farwell and Smith (2001) report was not a persuasive paper, as was discussed earlier. This publication contained no details on the dependent measures that would allow challenge of replicated methods; one simply could not subject the latter publication to CM challenge. Hence the focus on Farwell and Donchin (1991).

The Farwell and Donchin (1991) approach utilized

multiple probes, targets, and irrelevants as described above (the multiple was six). Within each subject, all stimulus types of a particular kind (e.g., all probes) are averaged into one bootstrapped average (see Rosenfeld et al., 2004 for a succinct description of the bootstrapping). BF does the same. Rosenfeld et al., (2004) called the Farwell-Donchin approach the six-stimulus protocol (6SP). The protocol of several reports from the Rosenfeld lab uses only one stimulus of each type (probe, target, irrelevant—but there are multiple irrelevants) per running block, and each stimulus is repeated several times; Rosenfeld et al., (2004) called it the one-stimulus protocol (1SP). Another difference between the protocols relates to the analysis type, but these differences are arbitrary; either protocol could use either analysis. In the 6SP and, one deduces, in BF, the cross correlation of probe and target waveforms is compared with the cross correlation of probe and irrelevant waveforms. (Rosenfeld et al., 2004 called this the FIT method.) Subjects with concealed knowledge produce P300s (or MERMERS) to probes because of recognition, and to targets because of instruction. So the probe-target "FIT" is expected to be bigger than the probe-irrelevant FIT (or cross correlation). In the 1SP, within each subject, one simply compares probe and irrelevant P300 wave amplitudes (SIZE method), which should differ in subjects with concealed information. (See Rosenfeld et al., 2004 for further details.)

In both protocols, our CM was to make the irrelevant stimuli relevant by having subjects execute covert mental or physical responses upon their appearance. That would make the probes and irrelevant P300s fit better and be the same size. To keep things simple, in the 6SP, Rosenfeld et al. (2004) asked subjects to learn one covert response per stimulus *category*. For example, there might be an operation-name category, consisting of animal names, as noted above. Subjects would execute one covert response to all irrelevants in that animal-name category. In the 1SP, each one (of say four) irrelevant items was assigned a particular covert CM response. These might involve imagining something, or unobservable toe wiggling, and so on.

The results were quite straightforward. Using the SIZE method in the 6SP, 82% of the "guilty" subjects who weren't taught CMs were correctly diagnosed. This figure was reduced to 18% in another group trained to use CMs. Using the preferred method of Farwell and Donchin, the FIT method, only 52 % of the simple guilty subjects were detected in the first place, and this was unchanged in the CM group—except these were not the same subjects detected. The FIT method thus appeared to work ran-

domly with the 6SP. Moreover, there was substantial overlap in the RT distributions for both simple guilty and CM subjects, so one couldn't detect CM use in any single individual case. Rosenfeld et al. (2004) discussed why the FIT method performed poorly in terms of out of phase probe and target responses, a problem which Farwell and Donchin (1991) didn't encounter with their paid volunteer subjects (versus our introductory-psychology pool subjects). Finally, and very importantly, there was nothing unusual about the wave averages in the CM group to the three stimulus types. The target P300 towered over the similarly sized probe and innocent P300s, just as in the innocent group, which we also ran. In summary, *CM users couldn't be detected*.

Things were just slightly better with our 1SP. That is, the SIZE method picked up 92% of the simple guilty subjects, but CM use by this same group one week later reduced this to 50%. (A habituation control group eliminated the alternative hypothesis that P300s to probes simply habituated over time.) However, the RT distributions showed no overlap between CM users and simple guilty subjects in response to irrelevant stimuli. Assuming that this result generalizes, one could in principle develop a criterion to detect CM use in the 1SP. Unfortunately, in a third week of testing with the subjects released from the CM requirement, 42% were still beating the test and looking perfectly innocent in terms of waveforms, and in terms of the total overlap of the RT distributions (which also showed that the CMs were indeed not being used, per the instructions to subjects). This presently unexplained and interesting fact could become the basis of a good CM strategy with the 1SP. The FIT method didn't fare terribly well here either, detecting only 69% of the simple-guilty subjects and only 25% of the CM users. In addition, Mertens et al. (2003), who, as was noted above, found very poor detection of guilty subjects in the first place—27% hit rate using the FIT method of analysis—reported that “CMs further lowered this rate of detection. . . . ERP procedures may be vulnerable to specific countermeasures”(page S25).

The preceding discussion casts doubt on the ability of P300-based recognition tests—including BF—to remain robust in the face of CMs. Yet there is another, very serious methodological consideration. From a theoretical perspective, the 6SP has difficulties not discussed prior to Rosenfeld et al. (2004). One might surmise that Farwell and Donchin (1991) chose to use six probes because the developer of the GKT (Lykken, 1981) used six items in his original study of the polygraph-based GKT. The point of using multiple items was as follows. If

for one item there is a choice of five evaluated alternatives, then the probability of a chance hit on that item is $1/5=0.2$. The use of more orthogonal items reduces the multiplied fractional chance-hit probabilities to, for example, 0.000064, with six items (.2 to the sixth power). With a six-item test, even hitting on just three items yields $p=.08$ chance-hit probability. (See also Ben-Shakhar & Furedy, 1990). The point is that in the format of a standard polygraph GKT, one has *separate responses to each, individual probe*. This is *not* the case with the six-probe paradigms of Farwell and Donchin (1991) or Farwell and Smith (2001), which average all probe P300s together. Let us suppose that an innocent subject produces a consistent P300 to just one and only one of the probes in a six-probe test—for whatever reason, such as actually recognizing this one guilty-knowledge item through press leakage. The resulting average ERP to all probes should contain a small P300, as it is an average of five actual irrelevants and one probe. The target will reliably produce a large P300. The FIT method, as Farwell and Donchin (1991) and BF use it, looks at cross correlations, which will, in calculating correlation coefficients based on standard scores, scale the amplitude differences between averaged probe and target away and likely declare guilt, not able to determine which or how many probe items were really recognized. The SIZE method might also find the probe greater than the irrelevant and also produce a false positive. This problem does not inhere in the single-probe protocol, which averages only identical stimuli. (See Rosenfeld et al., 2005b.)

CONCLUSIONS

One may read the following claim on the BF Web site:

Farwell Brain Fingerprinting is a revolutionary new technology for solving crimes, with a record of 100% accuracy in research with US government agencies and other applications. The technology is proprietary and patented. Farwell Brain Fingerprinting has had extensive media coverage around the world. The technology fulfills an urgent need for governments, law enforcement agencies, corporations, and individuals in a trillion-dollar worldwide market. The technology is fully developed and available for application in the field.

One might agree about the facts that the technology is proprietary and patented (indeed there are also several patents on P300-based detection of concealed information, the earliest of which antedate Farwell's), and that BF has had extensive media coverage, as noted above. There is considerable doubt, however, about fulfilling

urgent needs by U.S. government agencies. This may be confirmed by reading a report by the U.S. General Accounting Office to Senator Charles E. Grassley, titled "Federal Agency Views on the Potential Application of 'Brain Fingerprinting'" (U.S. General Accounting Office 2001). This report states that "Officials representing CIA, DOD, Secret Service, and FBI do not foresee using the Brain Fingerprinting technique for their operations. . . . CIA officials concluded that Brain Fingerprinting had limited application to CIA's operations. . . . Overall, DOD officials indicated that Brain Fingerprinting has limited applicability to DOD's operations. . . . According to FBI officials, the developer had not presented sufficient information to demonstrate the validity or the underlying scientific basis of his assertions. . . . the technique had limited applicability and usefulness to FBI. . . ." These summary views are elaborated in the complete report, which may be obtained on the Internet.

The last sentence from the BF Web page just quoted, "The technology is fully developed and available for application in the field," in the light of all the foregoing information, is, perhaps most charitably, viewed as florid advertising copy. This leads to a general question about this review. In writing a review of a technology, one usually relies on peer-reviewed, scientific literature, not Internet marketing material. In the case of BF, however, there is little choice, for, as noted, there has been only one serious publication that was based on a highly controlled laboratory situation with limited ecological validity—and it is vulnerable to CMs. Why bother reviewing advertising? The answer is that this Web material purports to be scientific, i.e., a technology is being marketed as a fully scientifically validated product. Thus, it seems fair to look at all the material and related documents on the Web site and investigate whether or not they are misleading. That question is left to the present reader or potential buyer, with the usual *caveat emptor*.

One should, however, conclude with the hope that the baby will not be thrown out with the bathwater: just because one person is attempting to commercialize brain-based deception-detection methods prior to completion of needed peer-reviewed research (with independent replication) does not imply that the several serious scientists who are now seriously pursuing this line of investigation should abandon their efforts. On the contrary, brain activity surely forms a substrate for deception which patient investigation may elucidate. It appears that detecting deception will continue to be of interest to various agencies and institutions. If it is to be done, it may as well be done well.

REFERENCES

- Allen, J., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology*, *29*, 504–522.
- Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception*. New York: Springer-Verlag.
- Ellwanger, J., Rosenfeld, J. P., Sweet, J. J., & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology*, *23*, 9–23.
- Ellwanger, J., Rosenfeld, J. P., Hannkin, L. B., & Sweet, J. J. (1999). P300 as an index of recognition in a standard and difficult match-to-sample test: A model of amnesia in normal adults. *The Clinical Neuropsychologist*, *13*, 100–108.
- Fabiani, M., Karis, D., & Donchin, E., (1983). P300 and memory: Individual differences in the von Restorff effect. *Psychophysiology*, *558* (abstract).
- Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related potentials. *Psychophysiology*, *28*, 531–547.
- Farwell, L. A., Martinerie, J. M., Bashore, T. R., Rapp, P. E., & Goddard, P. H. (1993). Optimal digital filters for long latency components of the event related brain potential. *Psychophysiology*, *30*, 306–315.
- Farwell, L. A. (1995). Method for Electroencephalographic Information Detection. *U.S. Patent No. 5,467,777*.
- Farwell, L. A., & Smith, S. S. (2001). Using brain MERMER testing to detect knowledge despite efforts to conceal. *Journal of Forensic Sciences* *46*(1), 135–143.
- Ford, C. V. (1996) *Lies! Lies!! Lies!!! The psychology of deceit*, Washington, D.C., American Psychiatric Press.
- Harrington, Terry J. v. State of Iowa*, November 14–15, 2000. No. PCCV073247.
- Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed information test. *Psychophysiology*, *33*, 84–92.
- Honts, C. R., & Amato, S. L. (2002). Countermeasures. In M. Kleiner, ed., *Handbook of polygraph testing*, New York: Academic Press, pp. 251–264.
- Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory*. New York: St. Martin's Press.
- Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist*, *35*, 409–420.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*, 385–388.
- Lykken, D. T. (1981). *A tremor in the blood*. New York: McGraw-Hill.

- Lykken, D. T. (1998). *A tremor in the blood*. Reading, Mass.: Perseus Books.
- Mertens, R., Allen, J., Culp, N., & Crawford, L. (2003). The detection of deception using event-related potentials in a highly realistic mock crime scenario. *Psychophysiology*, *40*, S60 (abstract).
- Miller, A. R., Rosenfeld, J. P., Soskins, M., Jhee, M. (2002). P300 amplitude and topography distinguish between honest performance and feigned amnesia in an autobiographical oddball task. *Journal of Psychophysiology*, *16*, 1–11.
- Miyake, Y., Mizutanti, M., & Yamahura, T. (1993). Event related potentials as an indicator of detecting information in field polygraph examinations, *Polygraph*, *22*, 131–149.
- Polich, J. (1999). P300 in clinical applications. In E. Niedermeyer and F. Lopes da Silva, eds. *Electroencephalography: Basic principles, clinical applications and related fields*, 4th Ed. (pp. 1073–1091). Baltimore-Munich: Urban & Schwarzenberg.
- Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, *38*, 319–335.
- Rosenfeld, J. P. (2002). Event-related potentials in the detection of deception, malingering, and false memories. In M. Kleiner, ed. *Handbook of polygraph testing*. New York: Academic Press, pp. 265–286.
- Rosenfeld, J. P., Nasman, V. T., Whalen, I., Cantwell, B., & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: A new method of lie detection. *International Journal of Neuroscience* *1987*, *34*, 125–129.
- Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *24*, 157–161.
- Rosenfeld, J. P., Ellwanger, J. W., Nolan, K., Wu, S., Bermann, R. G. & Sweet, J. J. (1999). P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model. *International Journal of Psychophysiology*, *33(1)*, 3–20.
- Rosenfeld, J. P., Rao, A., Soskins, M., & Miller, A. R. (2003). Scaled P300 scalp distribution correlates of verbal deception in an autobiographical oddball paradigm. *Journal of Psychophysiology*, *17*, 14–22.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205–219.
- Rosenfeld, J. P., Biroshak, J. R., and Furedy, J. J. (2005a). P-300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. In press, *International Journal of Psychophysiology*.
- Rosenfeld, J. P., Shue, E., & Singer, E. (2005b). Single versus multiple probe blocks of P300-based concealed information tests for autobiographical versus incidentally obtained information. In press, *International Journal of Psychophysiology*.
- Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). The case for peak-to-peak measurement of P300 recorded at .3 Hz high pass filter settings in detection of deception. *International Journal of Psychophysiology*, *40*, 173–180.
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatio-temporal analysis of the late ERP responses to deviant stimuli, *Psychophysiology*, *38*, 343–358.
- United States General Accounting Office (2001). *Investigative techniques: Federal agency views on the potential application of "Brain Fingerprinting."* GAO-02-2.

APPENDIX 1: IOWA SUPREME COURT DECISION EXCERPTS

Harrington, who was seventeen at the time, was charged with Schweer's murder and was ultimately convicted, primarily on the testimony of a juvenile accomplice, Kevin Hughes. . . . Hughes was impeached by the defense with prior statements he had made implicating other persons in the crime. Hughes had separately named three other men as the killer. Each man was ultimately discovered to have an alibi before Hughes finally fingered Harrington. Hughes admitted that he had also changed his testimony about the type of gun used, first stating that Harrington had a pistol, then a 20-gauge shotgun, and finally a 12-gauge shotgun. He conceded he was "a confessed liar," having lied "[a]bout five or six times talking about this case." Hughes acknowledged that he visited the murder scene with the police and prosecutor and told them what he thought they wanted to hear. At the time, Hughes was being held on various theft and burglary charges and "he was tired of [being in jail]." He admitted that these charges were dropped after he agreed to testify against Harrington and McGhee. . . . Harrington's claim under section 822.2(1) was based on an alleged due process violation arising from the prosecution's failure to turn over eight police reports to the defense during the criminal trial. See *Brady v. Maryland*, 373 U.S. 83, 87, 83 S. Ct. 1194, 1196–97, 10 L. Ed. 2d 215, 218 (1963) (holding failure of prosecution to disclose evidence that may be favorable to the accused is a violation of the Due Process Clause of the Fourteenth Amendment). . . . We also think the reports were "suppressed" within the meaning of the Brady rule. It is apparent from some of the questions asked by Harrington's defense counsel at trial that he had some information about a man seen walking a dog and carrying a shotgun near the railroad tracks by the car dealership. Gates is never mentioned by name, however, and

Harrington's first postconviction relief counsel testified that there were no police reports referring to Gates in the materials provided to him by the prosecutor in 1987. In addition, one of the lead investigators testified without impeachment at Harrington's 1988 PCR hearing that the police had no immediate suspects in the Schweer homicide. We think it probable that original trial counsel did not know that Gates was the suspicious person seen by witnesses in the area. Clearly, counsel did not know of Schweer's contact with a person fitting Gates' description in the nights preceding Schweer's murder, including the fact that Schweer caught this individual trying to break into a truck.

We conclude Harrington did not have the "essential facts" of the police reports so as to allow the defense to wholly take advantage of this evidence. . . . Upon our de novo review of the record and consideration of the totality of the circumstances, our collective confidence in the soundness of the defendant's conviction is significantly weakened. Hughes, the primary witness against Harrington, was by all accounts a liar and a perjurer. With the police offering a \$5000 reward for information, Hughes named three other individuals as the murderer before finally identifying Harrington as the perpetrator, and then only after the other three men produced alibis.

As questionable as Hughes' veracity is, it is not the character of the prosecution's principal witness that undermines our confidence in the defendant's trial; Hughes' ability and propensity to lie were well known in 1978. The unreliability of this witness is, however, important groundwork for our analysis because this circumstance makes it even more probable that the jury would have disregarded or at least doubted Hughes' account of the murder had there been a true alternative suspect. Gates was that alternative. See *Kyles*, 514 U.S. at 439, 115 S. Ct. at 1568, 131 L. Ed. 2d at 509 ("[T]he character of a piece of evidence as favorable will often turn on the context of the existing or potential evidentiary record.").

At the original trial Gates was one of more than a dozen individuals who were considered by the police as the potential culprit. Certainly defense counsel would not have had the time and resources to track down and investigate each of these individuals. But if the defendant had known the additional information contained in the withheld investigatory reports, the defense would surely have focused its efforts on Gates, not only in preparing for trial, but at trial as well. Our conclusion is based on two important points revealed in these reports: (1) Gates' identification as the suspicious person seen in the area with a gun and a dog; and (2) Schweer's contact

with Gates, which for the first time provided a concrete link between an alternative suspect and the victim.

The State is hard pressed to argue the defendant's trial preparation and trial strategy would not have been altered by this additional information. Officers testifying at the second PCR hearing admitted the police considered Gates to be "the prime suspect" based on their investigation, an investigation unknown to Harrington at the time of his criminal trial. It is fair to conclude that had Harrington's counsel been provided with this information, he would have zeroed in on Gates in his trial preparation and at trial, just as the police had zeroed in on Gates during their investigation. Harrington's attorney could have used Gates as the centerpiece of a consistent theme that the State was prosecuting the wrong person.

Independent witnesses placed Gates at the scene of the crime in the days before the murder. Independent witnesses saw him with a shotgun and a dog. The victim himself interrupted a person resembling Gates breaking into a truck only two nights before the victim was shot to death in the car lot. In contrast, Harrington was identified as the murderer by a confessed liar, whose testimony was corroborated only by two particles of gunpowder found on Harrington's coat several weeks after the murder and the now-recanted testimony of the witness's teenage cohorts. The murder weapon was never found and no one has ever connected Harrington with the dog prints found at the murder scene, even though the police from the beginning had focused their investigation on finding "a man with a dog."

Given this evidence, a jury might very well have a reasonable doubt that Harrington shot Schweer. That is all that is required to establish the materiality of the undisclosed evidence. See *Lay v. State*, 14 P.3d 1256, 1263 (Nev. 2000) (stating "specific evidence of the existence of another shooter" was potentially material because the defense "might develop reasonable doubt as to whether [the defendant] was the actual killer"). We do not think Harrington had to show, as the State argues, that the police reports would have "led to evidence that someone else committed [the] crime." It was incumbent on the State to prove Harrington's guilt beyond a reasonable doubt; it was not Harrington's responsibility to prove that someone else murdered Schweer. Therefore, if the withheld evidence would create such a doubt, it is material even if it would not convince the jury beyond a reasonable doubt that Gates was the killer.

Under the circumstances presented by the record before us, we cannot be confident that the result of Harrington's murder trial would have been the same had the exculpatory information been made available to him.

We hold, therefore, that Harrington's due process right to a fair trial was violated by the State's failure to produce the police reports documenting their investigation of an alternative suspect in Schweer's murder. See *Mazzan*, 993 P.2d at 74–75 (finding Brady violation where withheld "police reports provided support for [the defendant's] defense that someone else murdered" the victim); *Davis v. Commonwealth*, 491 S.E.2d 288, 293 (Va. Ct.

App. 1997) (holding prosecution's failure to disclose information of other African-American females in vicinity of drug sale constituted a Brady violation). Accordingly, we reverse the trial court's contrary ruling, and remand this matter for entry of an order vacating Harrington's conviction and granting him a new trial.

REVERSED AND REMANDED.