Scientific Consensus on Brain Fingerprinting and Differing Views on the Science, Technology, and Application

Revised October 23, 2011

The following proposed Scientific Consensus on Brain fingerprinting has arisen from discussions among forensic scientists, legal experts, psychophysiologists, and experts in law enforcement and national security. These discussions were initiated by Lawrence A. Farwell. This is a work in progress. Discussions of these and other related issues are ongoing. Please refer comments and suggestions to Lawrence A. Farwell at LFarwell@brainwavescience.com .

The most fundamental point of consensus among scientists and other relevant experts regarding brain fingerprinting, forensic science, and science in general is that different methods produce different results. Brain fingerprinting, from the seminal Farwell and Donchin (1986; 1991) and Farwell and Smith (2001) papers to the present, has never produced an error, neither a false negative nor a false positive. Some alternative methods of applying the same brain responses in attempts to detect concealed information have resulted in 10% to 15% errors and in some cases as high as nearly 50% errors, no better than chance. Even some purported “replications” of Farwell and Donchin have in fact used fundamentally different methods. Consequently they have failed to achieve accuracy approaching that of brain fingerprinting and, unlike brain fingerprinting, are susceptible to countermeasures. These fundamental differences in scientific methods are the reason why brain fingerprinting has been successfully applied in the field and ruled admissible in court, and these alternative methods are unsuitable for field use or application in the criminal justice system or national security.

In developing this consensus, we have specified precisely the standard scientific methods that constitute brain fingerprinting and attempted to identify the specific standards that are necessary and sufficient to obtain the results that brain fingerprinting has consistently attained. We have sought to identify differences in methods that are responsible for the widely divergent results obtained in different laboratories conducting related research.

Fundamental brain fingerprinting scientific principles, methods, and scientific standards are briefly described the first section of this article. The proposed Scientific Consensus on Brain Fingerprinting presumes a thorough understanding of the information contained therein. It also assumes familiarity with the articles in the literature cited in the Background section below.

In the course of developing a consensus, some points have arisen on which there is considerable diversity of opinion. Some of these Differing Views on Brain Fingerprinting are briefly outlined following the Scientific Consensus on Brain Fingerprinting.

Table of Contents

Background

The proposed Points of Consensus on Brain Fingerprinting presume familiarity with the following articles from the literature.

Representative peer reviewed articles on brain fingerprinting

Farwell, LA, & Donchin, E (1991). The truth will out: interrogative polygraphy (''lie detection'') with event-related potentials. *Psychophysiology 28*(5), 531–547. Available at: http://www.brainwavescience.com/Farwell_Donchin_1991_Psychophysiology_Brain_Fingerprinting_The_Truth_Will_Out.pdf

Farwell, LA & Smith, SS (2001). Using brain MERMER testing to detect concealed knowledge despite efforts to conceal. *Journal of Forensic Sciences 46*(1),135-143. Available at: http://www.brainwavescience.com/JourForensicScience.php

Farwell, L.A . (2011a). Brain Fingerprinting: Corrections to Rosenfeld. *Scientific Review of Mental Health Practice,* **8**, 2, 56-68.
Available at:
http://www.brainwavescience.com/Farwell_Brain_Fingerprinting_Corrections_to_Rosenfeld_Scientific_Review_of_Mental_Health_Practice.pdf

A more comprehensive version this article is cited as follows:

Farwell, L.A. (2011b). Brain Fingerprinting: Comprehensive Corrections to Rosenfeld in Scientific Review of Mental Health Practice.
http://www.brainwavescience.com/Scientific_Review_of_Mental_Health_Practice_Farwell_Brain_Fingerprinting_Comprehensive_Corrections_to_Rosenfeld.pdf . Seattle: Excalibur Scientific Press.

Available at:
http://www.brainwavescience.com/Scientific_Review_of_Mental_Health_Practice_Farwell_Brain_Fingerprinting_Comprehensive_Corrections_to_Rosenfeld.pdf.

Articles on legal issues

Farwell, L.A. and Makeig, T.H. (2005), Farwell Brain Fingerprinting in the case of *Harrington v. State. Open Court, X [10]:*3, 7-10. Indiana State Bar Association, September 2005. Available at:
http://www.brainwavescience.com/Farwell_Brain_Fingerprinting_in_Harrington_v_State

Roberts, A.J. (2007), Everything New Is Old Again: Brain Fingerprinting and Evidentiary Analogy, *Yale J. L. Tech 9*, 234-270. Available at: http://www.yjolt.org/files/roberts-9-YJOLT-234.pdf

Abstract (Psychophysiology 2011) on field and laboratory research on brain fingerprinting

Farwell L.A., Richardson D.C., Richardson G. (2011). Brain Fingerprinting Field Studies Comparing P300-Mermer and P300 ERPs in the Detection of Concealed Information. *Psychophysiology 48:* S96.

Abstract:
Brain Fingerprinting Field Studies Comparing P300-Mermer and P300 ERPs in the Detection of Concealed Information
Brain fingerprinting detects concealed information stored in the brain by measuring brainwave responses. We compared P300 and P300-MERMER event-related brain potentials (ERP) for accuracy and statistical confidence in four field/real life studies. Tests on 76 subjects detected presence or absence of information regarding 1) real crimes with substantial consequences (either a judicial outcome, including the death penalty or life in prison, or a $100,000 reward for beating the test); 2) real-life events including felony crimes; 3) knowledge unique to FBI agents; and 4) knowledge unique to explosives (EOD/IED) experts. With both P300 and P300-MERMER based analyses, determinations were 100% accurate: there were no false negatives, no false positives, and no indeterminates. Median statistical confidence for individual determinations was 99.9% with P300-MERMER and 99.6% with P300. Mean statistical confidence for individual determinations was 99.5% with P300-MERMER and 97.9% with P300. Countermeasures had no effect. An alternative, non-brain fingerprinting "complex trial protocol" had 0% accuracy and proved invalid, unreliable, and unusable in the field. All subjects figured out on their own how to beat the complex trial protocol. Brain fingerprinting accurately detected all of the same subjects. Scientific standards for brain fingerprinting research and field applications are discussed. All studies that have met these standards have achieved high accuracy. All studies that have reported low accuracy and/or susceptibility to countermeasures have failed to meet these standards.

Preprints available on request

Full article for the above abstract:
Farwell L.A., Richardson D.C., Richardson G. Brain Fingerprinting Field Studies Comparing P300-Mermer and P300 ERPs in the Detection of Concealed Information.

Comprehensive tutorial review on brain fingerprinting:
Farwell, LA. Brain fingerprinting: A comprehensive tutorial review of laboratory research and field applications. Evergreen Press. In press.

Brain Fingerprinting: Methods and Scientific Standards

Principles of Applying Brain Fingerprinting in the Field and the Laboratory

Generally there are three different processes involved in the application of brain fingerprinting science in a judicial case. These are 1) the investigation that precedes the science; 2) the objective, scientific procedure of brain fingerprinting; and 3) the weighing of the evidence, interpretation, and legal adjudication in terms of guilt or innocence that may follow later. Brain fingerprinting is an objective, scientific procedure. It does not depend on the subjective judgment of the scientist. It is preceded by a non-scientific process – investigation – and may be followed by another non-scientific process – judicial adjudication.

Before a brain fingerprinting test is conducted, a criminal investigator investigates the crime. He formulates an account of the salient features of the crime. These constitute the relevant knowledge to be tested in the brain fingerprinting test. (As described below, these constitute the probe stimuli used in the test, and also the target stimuli.) This criminal investigation is outside the realm of science. This process is based on the skill, expertise, and subjective judgment of the criminal investigator. The criminal investigator provides the scientist with the probe stimuli that in the criminal investigator's judgment represent the actual events involved in the crime.

The scientist applies the scientific procedure of brain fingerprinting to determine objectively whether or not the subject knows the crime-relevant information contained in the probes. Brain fingerprinting determines only the presence or absence of this specific information stored in the subject's brain. The brain fingerprinting scientist opines only on the presence or absence in the subject's brain of the specific knowledge embodied in the probes that were provided by the criminal investigator. Here the science ends.

The science and the scientist do not address the question of whether the results are probative of the subject's guilt or innocence, or whether the subject committed or did not commit any act. Brain fingerprinting science, and the brain fingerprinting scientist, do not even address whether the probes provided by the criminal investigator have anything to do with the crime. Brain fingerprinting does not evaluate whether or not the investigator's account of the crime is accurate, or whether the relevant knowledge correctly represents the crime, or whether any crime took place.

The judge and/or jury weigh the brain fingerprinting evidence, the criminal investigator's account of the crime, and all other evidence to reach a non-scientific, common-sense judgment regarding the suspect's participation in the crime. This process is outside the realm of science. They may reach a legal determination of guilty or not guilty. The role of the scientifically produced brain fingerprinting evidence is only to inform the trier of fact, not to render a scientific conclusion regarding guilt or participation in the crime.

A brain fingerprinting scientist can legitimately testify as an expert regarding only one specific fact: the subject does or does not know the relevant knowledge contained in the probes provided by the criminal investigator. The degree to which this fact is probative regarding the subject's participation in a crime is outside the realm of science, and outside the purview of the testimony of the brain fingerprinting scientist. That is a matter to be debated by the prosecution and defense and decided by a judge and/or jury.

Brain fingerprinting does not evaluate whether the subject should, could, or would know the information if he did or did not commit the crime or under any other real or hypothetical circumstances. It only determines whether or not the subject actually does know the relevant knowledge. The interpretation of the results of a brain fingerprinting test in terms of guilt or innocence, participation or non-participation in a crime, goes beyond the science and is outside the realm of expert testimony by a brain fingerprinting expert.

In short, brain fingerprinting is an objective, scientific process that is preceded by a process outside the realm of science and followed by another process outside the realm of science.

Brain fingerprinting is similar to DNA, fingerprints, and other forensic sciences in this regard. DNA, fingerprints, and all other forensic sciences also do not prove a subject guilty or innocent. The scientific data provided by a brain fingerprinting test – and the only subject on which a brain fingerprinting scientist testifies – are limited to a determination as to whether or not the information contained in the probes supplied by the criminal investigator is stored in the brain of the subject. Similarly, polymerase chain reaction (PCR) / short tandem repeats (STR) DNA testing determines only that a DNA sample putatively from the crime scene matches or does not match a DNA sample putatively from the subject. Like a brain fingerprinting test, a DNA test does not return a scientific outcome of guilty or a scientific determination that the suspect committed the crime. As discussed above, it is up to the judge and jury, not the scientist, to decide if brain fingerprinting evidence, taken along with all the other evidence, warrants a legal determination of guilty or not.

The purpose of brain fingerprinting is to determine whether or not specific relevant knowledge is stored in the brain of the subject.

In field cases, the relevant knowledge generally is information that an investigator thinks represent the details of a crime. Alternatively, it may be information that is known only to a particular group of people, such as FBI agents as in the FBI Agent Study reported herein, skilled bomb makers as in the Bomb Makers Study reported herein, trainees of an Al-Qaeda training camp, or members of a terrorist cell or hostile intelligence agency. In the CIA Real-Life Study and the Real Crimes Real Consequences $100,000 Reward Study, the relevant knowledge consists of information that the criminal investigator believes constitutes salient features of a crime that the perpetrator experienced in the course of committing the crime. The relevant knowledge is provided by the criminal investigator to the brain fingerprinting scientist. The goal of brain fingerprinting is to determine whether or not the relevant knowledge is known to the subject.

Unlike the present field studies, most previous brain fingerprinting tests have been conducted in laboratory settings. In a laboratory setting, the relevant knowledge is fabricated by the experimenter. One additional step is necessary before a test can be implemented to test whether or not the subject knows the relevant knowledge. The experimenter designs and implements a knowledge-imparting procedure to impart the relevant knowledge to the subject, as described below. The purpose of the knowledge-imparting procedure is to make certain that the subject knows the relevant knowledge. It generally consists of a training session and/or a mock crime.

The accuracy of a method to detect the relevant knowledge can only be evaluated when the relevant knowledge is actually there to be detected. To implement a valid study, it is necessary to determine independently that the knowledge-imparting procedure actually did impart the relevant knowledge to the subject, so the subject actually possesses the knowledge that the test is intended to detect. No detection technique, no matter how perfect, can detect something that is not there. Obviously, a scientific study to evaluate the effectiveness of a method to detect knowledge (or anything else) cannot be accomplished if the thing to be detected is not there. Post-test interviews are used to establish that knowledge-imparting procedure was successful and the information the test seeks to detect was actually there to be detected.

In a field case, the brain fingerprinting procedure begins after the criminal investigator has provided the relevant knowledge to the scientist. In a laboratory case, the brain fingerprinting procedure begins after the experimenter has fabricated the relevant knowledge and successfully implemented the knowledge-imparting procedure.

How the Brain Fingerprinting Test Works

In a brain fingerprinting test, stimuli are presented to the subject in the form of words, phrases, or pictures on a computer screen. Three types of stimuli are presented: targets, irrelevants, and probes.

Target stimuli are details about the investigated situation that the experimenter is certain the subject knows, whether or not he committed the crime. (We shall generally refer to the investigated situation as a "crime," although brain fingerprinting can of course be used to investigate non-criminal events as well.) The experimenter tells the subject about the target stimuli and their significance in terms of the crime. Because they are significant in the context of the crime to all subjects, targets elicit an "Aha" response in all subjects. Thus, targets elicit a corresponding P300-MERMER brain response whether the subject knows the other salient features of the crime or not.

Irrelevant stimuli contain information that is not relevant to the crime and not relevant to the subject. They consist of incorrect but plausible crime features. Irrelevant stimuli are designed to be indistinguishable from correct crime-relevant features to someone who does not know the features of the crime. Since the irrelevant stimuli are not significant in the present context, they do not elicit a P300-MERMER.

Thus, the targets and irrelevants both provide standard responses. The targets provide a standard for the subject's brain response to relevant, significant information about the crime in question. The irrelevants provide a standard for the subject's brain response – or rather lack of a response – to irrelevant information.

The third and most revealing type of stimuli is the probe stimuli. Probes contain information that is relevant to the crime or other investigated situation. Probes have three necessary attributes:

1. Probes contain features of the crime that in the judgment of the criminal investigator the perpetrators would have experienced in committing the crime;

2. Probes contain information that the subject has no way of knowing if he did not participate in the crime; and

3. Probes contain information that the subject claims not to know or to recognize as significant for any reason.

For example, if a subject claims not to have been at the murder scene and not to know what the murder weapon was, a probe stimulus could be the murder weapon, such as a knife. Irrelevant stimuli could be other plausible (but incorrect) murder weapons such as a pistol, a rifle, and a baseball bat.

For a subject who knows the relevant details about the crime, the probes, like the targets, are significant and relevant. Thus, the probes produce an "Aha" response when presented in the context of the crime. This manifests as a P300-MERMER in the brainwaves. For a subject who lacks the knowledge contained in the probes, the probes are indistinguishable from the irrelevants. Probes do not produce an "Aha" response or the corresponding P300-MERMER.

Subjects are instructed to press one button in response to targets, and another button in response to all other stimuli.

"All other stimuli" consist of probes and irrelevants. A subject who possesses the relevant knowledge recognizes the probes as a separate category. For a subject lacking the relevant knowledge, probes are indistinguishable from irrelevants. Such a subject does not recognize any difference between correct features of the crime (probes) and incorrect but equally plausible features of the crime (irrelevants).

The brain fingerprinting computerized data analysis algorithm computes a determination of "information present" or "information absent." The information that is either present or absent in the brain of the subject is the information contained in the probes. The brain fingerprinting system also computes a statistical confidence for each individual determination, e.g., "information present, 99.9% confidence." If there is insufficient data to reach either an "information present" or an "information absent" determination with a high statistical confidence, the algorithm returns the outcome of "indeterminate."

Note that an indeterminate result is not incorrect. It is not an error. It is neither a false negative nor a false positive. Rather, it is a determination that the data analysis algorithm has insufficient data to make a determination in either direction with a high statistical confidence.

Before conducting a brain fingerprinting test, the subject is interviewed to find out what he knows about the crime from any non-incriminating source such as news reports or prior interrogations. Any such information is excluded from the probes. (Such information may be contained in targets, since the targets are known to contain information that the subject knows.)

The experimenter describes to the subject the significance of each probe in the context of the crime. The experimenter does not tell the subject which stimulus is the probe and which are similar, irrelevant stimuli. Only information that the subject denies knowing is used for probe stimuli.

Also, the experimenter shows the subject a list of all the stimuli including the probes, without of course identifying which ones are probes. As an extra precaution, the subject is asked if any of the stimuli are significant to him for any reason at all. Any stimuli that are significant to the subject for reasons unrelated to the crime are eliminated. If for example, a potential probe is the name of a known accomplice, and coincidentally it is also the name of the suspect's brother-in-law, it is not used.

Things are significant to a person in context. The context of the probe stimuli in relation to the crime is established in the interview prior to the brain fingerprinting test. Immediately before the test, the experimenter describes the significance of each probe in the context of the crime. Before the test, the subject has explicitly stated that he does not know which stimulus is the probe containing the correct crime-relevant information.

Under these circumstances, a large P300-MERMER in response to the probes provides evidence that the subject recognizes the probes as significant in the context of the crime. If the scientist has followed the proper scientific protocols, the subject has eliminated all plausible non-incriminating explanations for this knowledge by his own account prior to the test. Therefore, an information-present response can provide evidence that a judge and jury may reasonably evaluate as being probative regarding the subject's involvement in the crime. Note, however, that the brain fingerprinting scientist does not opine on the subject of the suspect's guilt or participation in the crime.

The relevant knowledge is generated by the criminal investigator during the criminal investigation, prior to beginning the brain fingerprinting scientific procedures. The relevant knowledge generally comprises 12 to 30 short phrases or pictures, along with an explanation of the significance of each in the context of the crime. The investigator also provides the scientist with a detailed account of which items in the relevant knowledge are or may be already known to the subject for any known reason. For example, the investigator notes any specific features of the crime that have been published in the newspaper or revealed to the subject in interrogation or previous legal proceedings.

In field applications, before the scientific procedure of brain fingerprinting begins, a criminal investigator investigates the crime. This procedure is outside the realm of science, and depends on the skill and subjective judgment of the criminal investigator. The criminal investigator develops an account of the crime, including the salient features thereof. The criminal investigator provides the brain fingerprinting scientist with these salient features, which then constitute the probe stimuli to be tested in brain fingerprinting.

After the selection of the probe stimuli by the criminal investigator, the scientific procedure of brain fingerprinting begins. The objective, scientific procedure of brain fingerprinting comprises a test to determine whether the specific features of the crime contained in the probe stimuli provided by the criminal investigator are stored in the brain of a specific subject.

In field tests, the probe stimuli are provided by the criminal investigator, based on his investigation of the crime and his account of what took place.

In a laboratory study, the probe stimuli are simply made up by the experimenter. Before the subject can be tested in a laboratory setting, some kind of knowledge-imparting procedure must be undertaken to impart the knowledge of the probes to the subjects. Generally this is either some kind of mock crime, or a training procedure, or both. The purpose of the knowledge-imparting procedure is to ensure that the subject knows the information contained in the probes.

To be valid, laboratory studies must independently determine whether or not the knowledge-imparting procedure was effective in imparting the knowledge tested. This is accomplished by a post-test interview. The accuracy of a method to detect the relevant knowledge can only be evaluated when the relevant knowledge is actually there to be detected. If the knowledge-imparting procedure fails to impart the knowledge to the subject, then the knowledge is not there to be detected. Obviously, no technique, no matter how perfect, can detect knowledge that is not there.

In evaluating the accuracy of a method to detect concealed information in a laboratory situation, it is vital to determine independently that the knowledge-imparting procedure actually did impart the relevant knowledge to the subject, so the subject actually possessed the knowledge that the test was designed to detect. This is accomplished by post-test interviews.

If the knowledge-imparting procedure has failed to impart the knowledge, then a valid experiment on detection of that knowledge cannot be undertaken. If the experimenter does not determine through a post-test interview whether or not the knowledge-imparting procedure was successful in imparting the knowledge contained in the probes, then the results of the procedure to detect the knowledge are uninterpretable. In the case of an "information-absent" determination for a subject who engaged in the knowledge-imparting procedure, there is no way of knowing if the knowledge-imparting procedure failed to impart the knowledge and the knowledge-detection procedure correctly detected the resulting lack of knowledge (a true negative), or the knowledge-imparting procedure successfully imparted the knowledge and the knowledge-detection procedure failed to detect it (a false negative). In the absence of an independent measure of the effectiveness of the knowledge-imparting procedure, the knowledge-imparting procedure is confounded with the knowledge-detection procedure, rendering the study invalid and the results uninterpretable.

Since all of the present studies were field studies, we shall focus in our discussion on the procedures to be followed in the field, when the information to be tested consists of the features of real-life experiences encountered in the subject's actual life, outside the laboratory.

The relevant knowledge provided by the criminal investigator to the scientist generally contains six to nine or more items that have never been revealed to the subject. These constitute the probe stimuli. If there is an insufficient number of features that are known only to the perpetrator and investigators (probes), a brain fingerprinting test cannot be conducted.

Generally there are also six or more items that have already been revealed to the subject or are commonly known. These will constitute the target stimuli.

The test requires an equal number of targets and probes. If there are too few features already known to the subject, the experimenter may request additional information about the crime from the criminal investigator to use as target stimuli. Alternatively, if there are ample available features of the crime that are known only to the perpetrator and investigators (viable probes), the experimenter may elect to inform the subject about some of these features and use these as targets instead of probes.

Each stimulus presentation and the corresponding brainwave and behavioral (button-press) response is referred to as a "trial." Typically in a brain fingerprinting test, several groups of 70 – 100 trials are presented. Each group of trials is referred to as a "block." This is described in detail in the Methods section.

A necessary prerequisite for any test that can be applied in the real world is a behavioral task that requires the subject to read, process, and discriminate *every* stimulus, and to report on this discrimination through an overt behavioral measure *on each trial*. Otherwise, subjects could simply stare generally at the screen and know when each stimulus arrived, but not even read the stimuli that might produce incriminating brain responses. To meet this requirement, subjects are instructed to press a button (typically with one thumb) in response to target stimuli, and to press another button (with the opposite thumb) in response to all other stimuli. "All other stimuli" includes probes and irrelevants. The purpose of the button-press response is to ensure that the subject reads and processes every stimulus – most importantly, the probes – and proves that he has done so by an overt behavioral act on every trial.

In brain fingerprinting, every stimulus presentation or trial requires the subject to read, understand, and classify the stimulus. Subjects must push a different button in response to target stimuli than in response to other stimuli. Stimuli are presented in random order, so from the subject's perspective any upcoming stimulus might be a target. In order to determine whether or not the stimulus is a target, the subject must read every stimulus. If the stimulus is a target, the subject reads and processes the stimulus and presses the appropriate button. If the stimulus is a probe or an irrelevant, the subject reads and processes the stimulus and presses the other button. Subjects know that we record the accuracy of the button presses. Thus, they realize that they must read and process every stimulus in order to press the correct button for targets and nontargets (nontargets consisting of probes and irrelevants).

As discussed in detail in the Discussion section, alternative techniques that do not require the subject to read, process, and discriminate every stimulus and behaviorally report on this discrimination on each and every trial are not viable for field use where subjects may be covertly uncooperative.

We also record reaction time. Reaction time, however, is easily manipulated. Therefore it is not a viable measure for classifying and evaluating the status of subjects in real-world situations. For this reason, reaction times are not used in brain fingerprinting data analysis.

Data Analysis and Statistical Confidence in Brain Fingerprinting Tests

A brain fingerprinting test computes a determination of "information present" or "information absent" and a statistical confidence for this result. Recall that the target stimuli contain crime-relevant information that is known to the subject, whether or not he committed the crime. Targets provide a standard for the subject's brain response to information the subject knows and recognizes as significant in the context of the crime. These contain an "Aha!" response, a large P300-MERMER brainwave response. The responses to the irrelevant stimuli provide a standard for the subject's responses to information that is irrelevant or unknown. The irrelevant responses do not contain a large P300-MERMER brain response.

The purpose of data analysis in brain fingerprinting studies is to determine whether the probe responses, like the target responses, contain a telltale "Aha!" response characterized by a P300-MERMER brainwave pattern. Mathematically, this constitutes a procedure to determine whether the probe responses are more similar to the target responses or to the irrelevant responses. The procedure also provides a statistical confidence for this determination. To be of practical use, the procedure must compute a determination and statistical confidence for each individual subject.

To be valid, the statistical confidence for an individual determination of "information present" or "information absent" must take into account the level of variability in the individual brain responses that are aggregated in the average response. The statistical technique of bootstrapping computes a statistical confidence for each individual determination that takes this variability into account. Three publications Farwell (1992), Farwell and Donchin (1991), and Farwell and Smith (2001) describe this technique in detail. Wasserman and Bockenholt (1989) also included a description and analysis of Farwell and Donchin's application of bootstrapping in brain fingerprinting as an exemplar of value of this statistical technique in psychophysiology.

If the outcome of the bootstrapping procedure meets the predefined criterion for a high statistical confidence that the probe response is more similar to the target response than to the irrelevant response, then the determination is "information present." If the bootstrapping procedure meets the criterion for a high statistical confidence that the probe response is more similar to the irrelevant response, then the determination is "information absent."

If neither the statistical confidence for "information present" nor the confidence for "information absent" is high enough to meet established criteria, then the result is "indeterminate." Typically a confidence of 90% is required for an "information present" determination. A lower criterion, typically 70%, is generally required for an "information absent" determination.

Obviously, for the technique to valid, it is necessary to compute a statistical confidence not only for "information present" determinations but also for "information absent" determinations, and necessary to require a reasonably high statistical confidence for determinations in either direction.

Before applying the bootstrapping technique on correlations between waveforms, noise in the form of high-frequency activity is eliminated by the use of digital filters. Farwell and colleagues (Farwell, Martinerie, Bashore, Rapp, and Goddard 1993) have shown that a specific type of filters known as optimal digital filters are highly effective for eliminating this high-frequency noise while preserving the brainwave pattern of interest in event-related brain potential research. These filters are optimal in the precise mathematical definition of the word.

Scientific Standards for Brain Fingerprinting Tests

1. Use equipment and methods for stimulus presentation, data acquisition, and data recording that are within the standards for the field of cognitive psychophysiology and event-related brain potential research. These standards are well documented elsewhere. For example, the standard procedures Farwell introduced as evidence in the Harrington case were accepted by the court, the scientific journals, and the other expert witnesses in the case. Use a recording epoch long enough to include the full P300-MERMER. For pictorial stimuli or realistic word stimuli, use at least a 1800 millisecond recording epoch.

2. Use correct electrode placement. The P300 and P300-MERMER are universally known to be maximal at the midline parietal scalp site, Pz in the standard International 10-20 system.

3. Apply brain fingerprinting tests only when there is sufficient information that is known only to the perpetrator and investigators. Use a minimum of six probes and six targets.

4. Obtain the relevant knowledge from the criminal investigator (or for laboratory studies fabricate the relevant knowledge and implement an effective knowledge-imparting procedure). Divide the relevant knowledge into probe stimuli and target stimuli. Probe stimuli constitute information that has not been revealed to the subject. Target stimuli contain information that has been revealed to the subject after the crime.

5. If initially there are fewer targets than probes, create more targets. Ideally, this is done by seeking additional known information from the investigators. Note that targets may

contain information that has been publicly disclosed. Alternatively, some potential probe stimuli can be used as targets by disclosing to the subject the specific items and their significance in the context of the crime.

6. For each probe and each target, fabricate several stimuli of the same type that are unrelated to the crime. These become the irrelevant stimuli. For irrelevant stimuli, select items that would be equally plausible for a non-knowledgeable subject. The stimulus ratio is approximately one-sixth probes, one-sixth targets, and two-thirds irrelevants.
7. Ascertain that the probes contain information that the subject has no known way of knowing, other than participation in the crime. This information is provided by the investigator for field studies, and results from proper information control in laboratory studies.
8. Make certain that the subject understands the significance of the probes, and ascertain that the probes constitute only information that the subject denies knowing, as follows. Describe the significance of each probe to the subject. Show him the probe and the corresponding irrelevants, without revealing which is the probe. Ask the subject if he knows (for any non-crime-related reason) which stimulus in each group is crime-relevant. Describe the significance of each of the probes and targets that will appear in each test block immediately before the block, without naming the stimuli.
9. If a subject has knowledge of any probes for a reason unrelated to committing the crime, eliminate these from the stimulus set. This provides the subject with an opportunity to disclose any knowledge of the probes that he may have for any innocent reason previously unknown to the scientist. This will prevent any non-incriminating knowledge from being included in the test.
10. Ascertain that the subject knows the targets and their significance in the context of the crime. Show him a list of the targets. Describe the significance of each target to the subject.
11. Require an overt behavioral task that requires the subject to recognize and process every stimulus, specifically including the probe stimuli. Detect the resulting brain responses. Do not depend on detecting brain responses to assigned tasks that the subject can covertly avoid doing while performing the necessary overt responses.
12. Instruct the subjects to press one button in response to targets, and another button in response to all other stimuli. Do not instruct the subjects to "lie" or "tell the truth" in response to stimuli. Do not assign different behavioral responses or mental tasks for probe and irrelevant stimuli.
13. In order to obtain statistically robust results for each individual case, present a sufficient number of trials (stimulus presentations) of each type to obtain adequate signal-to-noise enhancement through signal averaging. Use robust signal-processing and noise-reduction techniques, including appropriate digital filters and artifact-detection algorithms. The number of trials required will vary depending on the complexity of the stimuli, and is generally more for a field case. Use an absolute minimum of 70 probe trials, preferably at least 100, and an equal number of targets. Present three to six unique probes in each block.

14. Use appropriate mathematical and statistical procedures to analyze the data. Do not classify the responses according to subjective judgments. Use statistical procedures properly and reasonably. At a minimum, do not classify subjects in a category where the statistics applied show that the classification is more likely than not to be incorrect.
15. Use a mathematical classification algorithm to classify the responses to the probe stimuli as being either more similar to the target responses or to the irrelevant responses. In a forensic setting, conduct two analyses: one using only the P300 (to be more certain to meet the standard of general acceptance in the scientific community), and one using the P300-MERMER (to provide the current state of the art).
16. Use a mathematical data-analysis algorithm that takes into account the variability across single trials.
17. Set a specific, reasonable statistical criterion for an information-present determination and a separate specific, reasonable statistical criterion for an information-absent determination. Classify results that do not meet either criterion as indeterminate. Recognize that indeterminate results are neither false positives nor false negatives. Report error rate directly (percentage equal to the number of false positives plus false negatives, divided by the number of information present plus information absent determinations), or report accuracy as 100% minus the error rate.
18. Restrict scientific conclusions to a determination as to whether or not a subject has the specific crime-relevant knowledge embodied in the probes stored in his brain. Recognize that brain fingerprinting detects only presence or absence of information – not guilt, honesty, lying, or any action or non-action. Do not offer scientific opinions on whether the subject is lying or whether he committed a crime or other act. Recognize that the question of guilt or innocence is a legal determination to be made by a judge and jury, not a scientific determination to be made by a scientist or computer.
19. Evaluate accuracy based on actual ground truth. As with any forensic science technique, ground truth is the true state of whatever the technique is designed to measure. For brain fingerprinting, ground truth is what information is stored in the subject's brain. Establish ground truth with certainty through post-test interviews in laboratory experiments and in field experiments wherein subjects are cooperative. Establish ground truth as accurately as possible through secondary means in real-life forensic applications with uncooperative subjects. Note that ground truth is what the subject actually knows, not what the experimenter thinks the subject should know. Ground truth is not what the subject has done or not done. Ground truth is not whether the subject is guilty, or deceptive.
20. Make scientific determinations based on brain responses. Do not attempt to make scientific determinations based on overt behavior that can be manipulated, such as reaction time.

Proposed Points of Consensus on Brain Fingerprinting

1. Brain fingerprinting detects concealed information that is known by a subject, or absence of same, by detecting the electroencephalographic signature of an information-processing brain response that is present if and only if the tested information is known.
    a. Brain fingerprinting does not detect whether a person is guilty of a crime. The question of whether the subject is guilty or not is a legal one to be decided by a judge and jury, not a scientific one to be decided by a scientist or computer.
    b. Brain fingerprinting scientists testifying as expert witnesses in court must confine their testimony to explaining the science and presenting data regarding whether a specific subject knows the information contained in specific probe stimuli. Questions of who did what and who is guilty or not guilty go beyond the science, and are the domain of the judge and jury.
    c. Scientists whose expert witness testimony on brain fingerprinting has been admitted in court have in the past confined their testimony to explaining the science and presenting data regarding whether a specific subject knows the information contained in specific probe stimuli. They have agreed that questions of who did what and who is guilty or not guilty go beyond the science, and are the domain of the judge and jury.
    d. Brain fingerprinting does not detect emotions, stress, intentions, or actions. It is based on cognitive information processing in the brain.
    e. Brain fingerprinting does not detect lies. Any forensic science technology, including brain fingerprinting, can be used to indirectly reveal a lie. (For example, a subject may state that he knows nothing about a crime, and brain fingerprinting may demonstrate that he has relevant knowledge and therefore his previous statement is a lie; similarly, he may state that his DNA could not possibly match DNA from a crime scene, and DNA science could demonstrate indirectly that this was a lie.) The results of a brain fingerprinting test will be the same whether the subject previously has told the truth, lied, or not spoken about his knowledge of and/or participation in a crime.
2. To construct a valid brain fingerprinting test, it is necessary to use probes containing information that the subject has not been exposed to after the crime and that the subject denies knowing or recognizing as significant. Without such information, a brain fingerprinting test cannot be structured.
3. Brain fingerprinting data analysis can be and has been conducted in at least two ways: using the positive peak of the P300 alone, and using the positive P300 peak followed by the late negative peak (LNP). (The two together have been referred to as the P300-MERMER.)
    a. Analysis that includes both the positive P300 peak and the later negative deflection has generally produced more accurate results than analysis with the P300 alone.

   b. Differences in terminology exist. Over 1,000 publications have use the term "P300" to refer to only the positive peak; a handful have used the term "P300" to refer to the positive peak followed by the negative peak, i.e., they have defined P300 amplitude as peak-to-post-peak. In other words, a handful of authors have used the term "P300" to refer to what brain fingerprinting scientists refer to as "P300-MERMER." God has not yet weighed in on which is the more correct term.
   c. The additional attributes of the P300-MERMER, in addition to the positive P300 peak followed by the late negative peak (LNP), are a field for future research.

4. Bootstrapping is a statistical method to compute the probability that particular data have particular characteristics, a method that makes no assumptions regarding the distribution of the data and therefore is highly tolerant of different distributions.
   a. Bootstrapping has the advantage of allowing for comparison between average responses, with the concomitant signal-to-noise enhancement inherent in signal averaging, while taking into account the variability across the single trial brain responses that make up the averages.
   b. Bootstrapping has been effectively used to compute the probability that the probe response is more similar to the target response than to the irrelevant response. If this probability is greater than a criterion (usually 90%), then the determination is "information present." 100% minus this probability is the probability that probe response is more similar to the irrelevant response than to the target response. If this is higher than a second criterion (usually 70%), then the determination is "information absent." If neither criterion is met, no determination is made, and the result is "indeterminate." (Indeterminate results are neither a false positive nor a false negative; they are not an error.)
   c. Correlations have been used with bootstrapping as a measure of similarity between probes-targets and probes-irrelevants. Amplitude and area can also be used.
   d. Greater accuracy and validity are obtained by using reasonable criteria for "information present" and "information absent" determinations. At a minimum, it is not valid or reasonable to classify response data in a category where the bootstrapping or other statistical method used computes that there is greater than a 50% probability that the chosen classification is incorrect (i.e., classifying a subject's data as "information absent" when there is greater than a 50% probability that the correct classification is "information present;" see Appendix 2.)

5. Studies that have met all 20 of the brain fingerprinting scientific standards have resulted in no false positives and no false negatives under any conditions, in the laboratory or in the field. (If anyone knows of an exception to this, please inform.)
   a. It is correct to state, regarding a specific set of data wherein there were no classification errors, that the results in that particular experiment or series of

experiments constituted 100% accuracy for the specific methods applied, but such a statement must be restricted to specific past results already obtained.

   b. It is incorrect to characterize any science or technology as “100% accurate” in general, because such a general characterization inherently includes a prediction that the technology will produce no errors in the future, which cannot be known.

6. Studies that have met some but not all of the 20 brain fingerprinting standards have resulted in varying accuracy rates, depending on the methods applied. Error rates for different methods have ranged from no better than chance accuracy to less than 10% false positive/negative errors.
7. Studies that have met some but not all of the 20 brain fingerprinting standards have resulted in varying results regarding countermeasures. In some studies countermeasures reduced accuracy considerably, in some they did not.
8. Although the available results suggest that meeting all of the 20 brain fingerprinting standards may be sufficient condition to achieve valid, reliable, and accurate results, it has not been shown that all of the 20 brain fingerprinting standards are necessary conditions. In some cases valid experiments have been conducted that have resulted in relatively high accuracy without meeting all of the brain fingerprinting standards.
9. The available data suggest that some but not all of the brain fingerprinting standards are necessary conditions for validity, accuracy, and/or reliability.
   a. Standards 1; 2; 3 (first part); 4 (first part); 6; 7; 9; 11; 12; most of 13; 14; parts of 15; 16; 17; 18; 19; and 20 are necessary conditions.
   b. Standard 3 (second part) is not necessary; fewer probes can be used, although this may reduce accuracy.
   c. Standard 4 (second part), 5, and 10 are not necessary, in the following sense. Instead of using targets that are relevant to the crime and also disclosed to the subject, targets may be irrelevant but made relevant by task instruction to press a particular button in response to targets and another button in response to all other stimuli.
   d. Standard 8 is not necessary. Even if the significance of the probes is not stated in experimental instructions, the subject may recognize the probes as significant and emit the expected brain response. Electing not to explain the significance of the probes, however, may reduce accuracy. Not finding out if the probes are significant to the subject for a non-crime-related reason may create ambiguities in interpreting the data. Not describing the significance of the probes and targets immediately before each block may reduce accuracy.
   e. Parts of standard 13 are not necessary. A test can be successfully run with fewer than 3 or more than 6 unique probes per block, although this may reduce accuracy in real-world applications.
   f. Parts of standard 15 are not necessary. It is not necessary to conduct two separate analyses, one with the P300 and one with the P300-MERMER. Either one will suffice, although conducting both has advantages as stated in the standards.

10. Brain fingerprinting field tests are preceded by a criminal investigation that is outside the realm of science, and may be followed by a process of legal adjudication that is also outside the realm of science and is conducted by the judge and jury based on their common sense and human judgment.
    a. Prior to a brain fingerprinting field test, a criminal investigator investigates the crime and formulates his account of the crime. This process is outside the realm of science, and is based on the criminal investigator's skill and judgment.
    b. Brain fingerprinting does not determine whether or not the criminal investigator's account of the crime and the probes he provides based thereon accurately represent the crime, whether a crime even took place, or whether any individual participated in the crime.
    c. The scientific procedure of brain fingerprinting only determines if the subject knows the information contained in a specific set of probe stimuli and recognizes it as significant in the context of the crime.
    d. The judge and jury take into account the brain fingerprinting evidence along with all other available evidence, and reach their determinations regarding who participated in the crime and who is guilty or not based on their human judgment and common sense.
    e. Brain fingerprinting provides an objective account of the contents of human memory. Witness testimony provides a subjective (and not always truthful) account of the contents of human memory. In extrapolating from the contents of memory as revealed by witness testimony or brain fingerprinting, judges and juries must take into account the well known limitations of human memory. Human memory is not perfect: it is influenced by mental and physical illness, injury, passage of time, aging, trauma, drugs, and many other well known factors. Judges and juries must take into account these well known limitations in any trial that involves brain fingerprinting, just as they must in any trial that involves witness testimony.
11. Regarding the Harrington case
    a. The District Court ruled as follows:
        i. "In the spring of 2000, Harrington was given a test by Dr. Lawrence Farwell. The test is based on a 'P300 effect'."
        ii. "The P-300 effect has been recognized for nearly twenty years."
        iii. "The P-300 effect has been subject to testing and peer review in the scientific community."
        iv. "The consensus in the community of psycho-physiologists is that the P-300 effect is valid."
        v. "The evidence resulting from Harrington's 'brain fingerprinting' test was discovered after the verdict. It is newly discovered."

b. Although the District Court judge ruled brain fingerprinting science and Farwell's and Iacono's testimony based thereon admissible, he stopped short of granting Harrington a new trial, stating that the newly discovered evidence was insufficient to have probably changed the original verdict. Harrington appealed to the Supreme Court of Iowa, which overturned his conviction based on a constitutional rights violation in the initial trial. The Supreme Court did not reach the brain fingerprinting issue, and let the law of the case stand regarding the admissibility of brain fingerprinting.

12. Regarding Miyake et al.'s (1993) study in Japan.

    a. The authors cited the seminal Farwell and Donchin (1991) paper on brain fingerprinting.

    b. Miyake et al. failed to meet 18 of the 20 brain fingerprinting standards (all but numbers 3 and 20).

    c. They failed to implement data collection, artifact rejection, and data analysis procedures that meet the universal standards met by other laboratories in the field of event-related brain potential research, as follows. They measured P300 from the wrong scalp location. Their classifications were based not on any mathematical algorithm but on subjective judgments by the operators. They failed to use well-known standard methods, or any method, for artifact rejection or correction, resulting in inadequate data for accurate analysis or conclusions. Their timing parameters were outside the range used in other laboratories in event-related potential research. They used an insufficient number of trials. They attempted to detect lying, rather than information.

    d. These errors resulted in an exceptionally low accuracy rate. Only 65% of their determinations were correct, with 17% indeterminate.

    e. Their results are not an accurate or valid representation of the accuracy of brain fingerprinting or other techniques that use adequate scientific methods in the detection of concealed information

13. Studies in which brain fingerprinting detects real-life information regarding actual crimes, with the concomitant complications and motivations inherent thereto, in which the results of the test are potentially life changing due to judicial consequences such as the death penalty or life in prison – or to substantial rewards such as $100,000 cash for beating the test – may better reflect the validity, reliability, and accuracy of the techniques tested than laboratory studies in which there are no non-trivial consequences.

14. Brain fingerprinting has been published in the leading journals in both psychophysiology (*Psychophysiology*, Farwell and Donchin 1991) and forensic science (*Journal of Forensic Sciences,* Farwell and Smith 2001).

    a. Both of these articles meet the brain fingerprinting scientific standards.

    b. Both Donchin and Farwell have said (under oath in expert witness testimony) that Farwell and Donchin 1991 presents the same technique that Farwell later called brain fingerprinting.

c. Both Smith and Farwell have said the same thing about the Farwell and Smith 2001 paper.

d. *Psychophysiology* is the official peer-reviewed journal of the Society for Psychophysiological Research, and is recognized among psychophysiologists as a leading psychophysiology journal.

e. *The Journal of Forensic Sciences* is the official peer-reviewed journal of the American Academy of Forensic Sciences, and is recognized among forensic scientists as one of the leading journals in the field. Its impact factor of 1.524 is among the highest for forensic science journals.

15. Brain fingerprinting may be considered a kind of guilty knowledge test (GKT) or concealed information test (CIT). Brain fingerprinting also has some differences from the conventional GKT as applied with peripheral autonomic-nervous-system-based physiological measures. It measures cognitive information processing in the brain, rather than physiological arousal. Brain fingerprinting includes target stimuli, which provide a standard for known information, as well as probes (relevant stimuli) and irrelevant stimuli. The conventional GKT involves only relevant and irrelevant stimuli.

16. Statistics must compute a determination of "information present" or "information absent" (by whatever name) and a statistical confidence for each individual determination. To be classified as information present, data analysis must return a high statistical confidence that the subject is in fact information present. To be classified as information absent, data analysis must return a high statistical confidence that the subject is in fact information absent. Reasonable criteria must be established for both classifications. Outcomes, if any, where data fail to meet a reasonably high statistical confidence criterion for either information present or information absent are in fact indeterminate, and must be correctly reported as such. Obviously, subjects must not be classified in a category (information present or absent) where the statistics computed determine that the probability is less than 50% that the correct determination is the selected category (or there is a greater than 50% probability that the correct determination is the *other* category).

17. For any forensic science test where there are potential judicial consequences, error rates must be clearly stated. They should also be clearly stated in any laboratory study of a technique that may potentially have judicial consequences. Judicial proceedings, and specifically the Daubert standard, require correct reporting of error rates. Any method of reporting that disguises or conceals the actual error rate is unacceptable. True positives, false positives, true negatives, false negatives, and indeterminates must be clearly distinguished. The error rate is the percentage of determinations made that are either false positive errors or false negative errors. Indeterminates, if any, must also be clearly stated and correctly identified. Indeterminates are neither false positive errors nor false negative errors: they are not errors. If error rates are not stated directly as error rates but rather are stated indirectly in terms of accuracy rates, accuracy rates must be stated in such a way that actual error rates can be readily determined: the error rate is 100% minus the accuracy rate. "Accuracy" rates that fail to distinguish between false negative/false positive errors on the one hand and indeterminates on the other hand are unacceptable because they make it impossible to determine the actual error rate. This makes

comparison between studies impossible or misleading, and also disallows meaningful application in the judicial system.

18. Studies have compared P300 and P300-MERMER event-related brain potentials (ERP) for accuracy and statistical confidence in field/real life studies including, among others, tests on 76 subjects that detected presence or absence of information regarding the following:

    a. real crimes with substantial consequences (either a judicial outcome, including the death penalty or life in prison, or a $100,000 reward for beating the test);

    b. real-life events including felony crimes;

    c. knowledge unique to FBI agents; and

    d. knowledge unique to explosives (EOD/IED) experts.

    With both P300 and P300-MERMER based analyses, error rate was 0%. Determinations were 100% accurate: there were no false negatives, no false positives, and no indeterminates. Median statistical confidence for individual determinations was 99.9% with P300-MERMER and 99.6% with P300. Mean statistical confidence for individual determinations was 99.5% with P300-MERMER and 97.9% with P300. Subjects were taught the same countermeasures that have proven effective against other, non-brain fingerprinting techniques. Countermeasures had no effect on brain fingerprinting accuracy.

Differing Views on Brain Fingerprinting and Related Science and Technology

1. Although brain fingerprinting has never resulted in any false positives or false negatives – 100% of determinations made have been correct – other studies that were based on the three-stimulus P300-based paradigm introduced by Farwell and Donchin but which have failed to meet a substantial portion of the brain fingerprinting scientific standards have reported substantially lower accuracy rates than brain fingerprinting, in some cases no better than chance. Studies of methods that fail to meet even half of the brain fingerprinting standards have been found to be susceptible to various countermeasures.
    a. Brain fingerprinting experts attribute the inaccuracy and susceptibility to countermeasures of other studies to the failure of these alternative methods to meet the brain fingerprinting standards, in other words, on the *differences* between these alternative techniques and brain fingerprinting. In their opinion, the inaccuracy and the susceptibility to countermeasures of other techniques does not reflect negatively on brain fingerprinting.
    b. Some others attribute the inaccuracy and susceptibility to countermeasures of other, non-brain fingerprinting techniques to *similarities* between these other studies and brain fingerprinting, rather than to differences between the other studies and brain fingerprinting. Some authors have opined that the reported inaccuracy and susceptibility to countermeasures of other non-brain fingerprinting techniques implies that brain fingerprinting also must be inaccurate and susceptible to countermeasures. Some purported replications of Farwell and Donchin have in fact met fewer than half of the brain fingerprinting scientific standards.
2. There is considerable diversity of opinion regarding whether the P300 is a single phenomenon, or is comprised of a number of components, and, if there is more than one component, the attributes of the various subcomponents. There is also diversity of opinion on whether the P300-MERMER is a unified phenomenon, or whether it is composed of a number of separate components or subcomponents. There is diversity of opinion on what the P300-MERMER includes (or should include) in addition to the positive P300 peak and the late negative peak (LNP). No amount of words is likely to resolve any of this at present. All of this is to be resolved by future research.
3. There is considerable diversity of opinion on whether the Iowa District court ruled correctly in admitting brain fingerprinting as scientific evidence. There is diversity of opinion on whether brain fingerprinting should be admitted as evidence in the future under the Daubert standard and/or the Frye standard.
4. Farwell defines a brain fingerprinting expert is as a scientist who meets or exceeds the following standard: An individual who has published brain fingerprinting research in peer-reviewed psychophysiology journals and also in peer-reviewed forensic science journals; who has successfully applied brain fingerprinting in addressing real-world criminal cases in the field, with the concomitant complications and motivations; who also

has conducted rigorous and successful laboratory research on brain fingerprinting; who has testified as an expert witness in court; whose research and field applications have consistently met the brain fingerprinting scientific standards outlined in the attached paper; and whose accuracy rate in all research and applications has been high enough for practical field use (in Farwell's opinion, that means ideally less than 1% false negatives/positives, or in any case less than 1% false positive/negative errors overall and less than 5% false negatives/positives in each and every study and all field applications). Differing views exist. Others consider some scientists with lesser qualifications and achievements to be experts as well.

5. There is considerable diversity of opinion on what constitutes sufficient accuracy for a technique to be viable for field use. Farwell's opinion is that a technique that is viable for field use ideally should in every laboratory study and every series of field applications have produced less than 1% false negatives/positives, or in any case should have produced less than 1% false positive/negative error rates overall and less than 5% false negative/positive errors in each individual study. Others consider higher error rates and less consistent performance to be hypothetically acceptable for field use. Other than Farwell and colleagues, however, no other researchers in the USA have attempted to use their alternative, non-brain fingerprinting techniques in any field applications or any real-world situations with non-trivial consequences.

6. There is substantial consensus, in fact probably complete unanimity, among scientists regarding what brain fingerprinting measures. It detects the presence or absence of specific information or knowledge by detecting specific information-processing brain activity that takes place if and only if the specific knowledge tested is present. There are differing opinions on what it *should* measure.

    a. Farwell's opinion is that brain fingerprinting should measure what it is designed to measure and what it does measure as stated immediately above.

    b. Some have expressed the opinion that brain fingerprinting *should* measure various things that it is not designed to measure and does not measure: Some have expressed the opinion that it should measure guilt or innocence, and the fact that it does not is therefore a weakness; some have expressed the opinion that it should detect participation in past acts, and the fact that it does not is therefore a weakness; some have expressed the opinion that it should detect intention, and the fact that it does not is therefore a weakness; some have expressed the opinion that it should detect lies, and the fact that it does not is therefore a weakness; some have expressed the opinion that it should detect information that an individual does not know, but that is related to something that the individual has done, and the fact that it does not is therefore a weakness; some have expressed the opinion that it should detect what an individual should know, could know, or would know under certain circumstances (e.g., if he committed a crime), and the fact that it detects specifically what an individual actually does know is therefore a weakness; some have expressed the opinion that it should detect past actions, and the fact that it detects knowledge rather than actions therefore is a weakness.

7. There is diversity of opinion on what would be the best term to name brain fingerprinting. Some think that the term is apt; others do not.

8. There exists an unfortunate and confusing diversity of terminology for reporting error rates. In a judicial context, the important figure is the error rate (for example, rate of error is specified in the Daubert standard for admissibility). In some lay usage, "accuracy" is more commonly discussed. Ambiguity can arise when terms are not consistently used and defined. Different authors use the term "accuracy" to mean different things. Standard reporting for brain fingerprinting accuracy rates and error rates is as follows. Each outcome is either a true positive, a true negative, a false positive, a false negative, or an indeterminate. Error rates are represented thus: "In the cases where a determination was made, x% were errors: the error rate was x%. Of the information present subjects, y% were false negatives. Of the information absent subjects, z% were false positives. w% of the outcomes were indeterminate (not an error)." Accuracy rates are represented thus: "x% of the tests resulted in a determination with a high statistical confidence of either information present or information absent. y% of the determinations made were correct: the accuracy rate was y%; z% of the information present determinations were correct; w% of the information absent determinations were correct."

   Some authors report "accuracy" rates without distinguishing between indeterminates and errors. This is not a viable method for reporting accuracy rates for a number of reasons. The most important reason is that it conceals the false positive and false negative error rate, thus making comparison with other studies and consideration of the legal implications of the error rate impossible. Consider, for example, the terminology when applied to a study like Farwell and Donchin (1991) which had 12.5% indeterminate, and 100% correct determinations – no false positives and no false negatives. Consider another (hypothetical) study that had no indeterminates, and 12.5% errors (12.5% false negatives, and 12.5% false positives). With standard error-rate terminology as is used in brain fingerprinting and universally used in the criminal justice system, Farwell and Donchin had 0% errors and 12.5% indeterminates. The hypothetical study had 12.5% errors. These are considerably different outcomes, particularly if this were field cases and one happens to be one of the subjects for whom a false positive error is made. Expressed in standard terms of accuracy, the studies would be expressed thus: Farwell and Donchin reached a definite determination in 87.5% of cases; 100% of determinations were correct; error rate was 0%. The other study would be represented thus: All cases returned a definite determination, 87.5% of determinations were correct; error rate was 12.5%. Again, the actual differences in the data are accurately represented in the description of the results.

   Consider, on the other hand, the non-standard "accuracy" reporting that ignores the difference between indeterminates and errors, and thus conceals the actual rate of false positive and false negative errors. The data would be reported thus: Farwell and Donchin had 87.5% "accuracy." The other study had 87.5% "accuracy" also. This is misleading, because in the Farwell and Donchin study, there was no chance that a particular outcome was a false positive or false negative error. In the other study, there was a significant chance of an error – namely 12.5%. Again, this distinction is critical in a real-world situation where an error can mean life or death, freedom or imprisonment.

In a forensic situation, the critical question is, given that the outcome of a test is information present or information absent, what is the probability that this is an error? In the real world of criminal justice, a probability of 0% or near 0% is very different from a probability of 12.5%. Whatever may be acceptable in the laboratory, in the real world a 12.5% probability of being falsely subject to the death penalty or life in prison is very different from a 0% probability of being falsely subject to the death penalty or life in prison.

In standard reporting in forensic science, all error reporting and "accuracy" reporting must state the actual results, that is, must clearly state false positive and false negative errors on the one hand and indeterminates (if any) on the other hand. Some publications on detection of concealed information written by authors without real-world experience in forensic science, however, have failed to meet this standard. In real-world forensic science, "accuracy" means percentage of determinations that are correct. This use of the term is necessary in order for to make it possible from "accuracy" figures to compute the actual error rate, which is central to judicial proceedings and practical implications of the outcome in the real world. The percentage of indeterminates (or of determinations) must of course be clearly stated as well.

In Farwell's view, it is important to avoid shortcuts in reporting figures that fail to distinguish between categories of outcome that are in reality extremely different in scientific meaning and real-world consequences. "Accuracy" rates that conceal the actual false positive and false negative error rates by combining these with indeterminates are not viable for use in the criminal justice system. This is discussed in more detail in Appendix 1, "Correct and Accurate Reporting of Error Rates in Forensic Science." Some others are of the opinion that reports that confound false positive and false negative errors with outcomes in which no determination was made, and thereby conceal the actual false negative and false positive error rates, are acceptable in the laboratory. No one, however, has attempted to use any such technique in the field, as this clearly would not be valid or legally defensible, and could lead to gross miscarriages of justice and violations of human rights.

There are a number of other points of diversity of opinion, some major and many very minor.

Appendix 1
Correct, Consistent, and Accurate Error Reporting in Forensic Science

For judicial purposes, the figure of primary importance in science is the error rate. The error rate figures prominently in the evaluation of scientific evidence under the prevailing Daubert standard.

The importance of the error rate is that it provides a method of estimating the probability that a particular determination that has been made is in fact an error, based on the variability in the population. The law recognizes that as a practical matter what is important in an individual case is this: When a scientific procedure has returned an outcome with potential judicial consequences – in this case, an "information-present" or "information-absent" determination – what is the probability that this outcome is an error?

Indeterminates do not figure in the calculation of the probability that a given definite outcome is an error, because they are neither false positive nor false negative errors. When a determination is made in a specific case, the number of indeterminates that have taken place previously does not increase or decrease the probability that the determination made in the present case is an error.

Moreover, as a practical matter, an indeterminate does not provide evidence that the subject knows the information tested, nor does it provide evidence that the subject does not know the information.

Due to the central importance of error rates in judicial proceedings, it is essential in reporting results to report the error rate directly, or, alternatively, to report the accuracy rate correctly – that is, as 100% minus the error rate – so that the error rate can readily be determined.

It is not valid to report "accuracy" as representing the correct determinations as a percentage of the total tests run, because this confounds indeterminates – which are not an error – with false positives and false negative errors. A report of "85% accuracy" which confounds indeterminates with false positives and false negatives makes it impossible for the reader to determine whether the technique had an error rate of 15% with no indeterminates, or an error rate of 0% with 15% indeterminates, or something in between. An error rate of 15% is obviously very different from an error rate of 0%, particularly when there are real-world consequences to the outcome. A technique with a historical error rate of 0% is clearly accurate enough for field use, whether there are indeterminates or not. A technique with an error rate of 15% is not nearly accurate enough for field use. Failure to distinguish between indeterminates on the one hand and false positive/negative errors on the other results in a reporting method that hides the actual error rate and makes meaningful comparison with studies that correctly report the accuracy and error rate impossible.

In a forensic situation, the critical question is, given that the outcome of a test is information present or information absent, what is the probability that this is an error? In the real world of criminal justice, a probability of 0% or near 0% is very different from a probability of 15%. Whatever may be acceptable in the laboratory, in the real world a 15% probability of being falsely subject to the death penalty or life in prison is very different from a 0% probability of being falsely subject to the death penalty or life in prison. Any technique that disguises the actual false positive / false negative error rate by confounding false positive / negative errors with outcomes in which no determination is made (indeterminates) is unsuitable for field use. Fortunately, no one has attempted to use such techniques in the field, although they have been used in the laboratory. Use of such techniques in the field would inevitably result in gross miscarriages of justice as well as human rights violations.

Applicability must not be confounded with accuracy. In every forensic science, there is the possibility of false positive and false negative errors, and there are also cases where the data are insufficient to make a determination. For example, there may be some residue left by fingers at a crime scene, but the prints are so smudged that it is impossible to determine whether or not they match a known print. Biological samples may be so degraded that is not possible to determine a DNA match or non-match with a high confidence. In these cases, the forensic science does not make a false negative or false positive error. The technique is not applicable. The result is indeterminate. The percentage of cases where a determination is made must not be confounded with the percentage of determinations that are correct. The former is the applicability rate, and the latter is the accuracy rate.

Correct ways to report an outcome of 15% indeterminates and 0% errors include "100% accurate and 85% applicable," "0% error rate with 15% indeterminates," and "100% accurate with 15% indeterminates." Obviously, when there are indeterminates, they must be reported fully, consistently, and accurately. (When using statistics that do not allow for an indeterminate outcome, the rate of applicability is always 100% and need not be separately stated with each result, as long as it is clearly stated that indeterminates do not occur because the statistics do not allow them.)

When generalizing from particular results to a prediction regarding the population, the number of cases in which a determination has been made is also important. If 0 errors have been made in 100 determinations of information present or information absent, the expected value of the error rate for the population is more accurately represented as "less than 1%" than as "0%." If 0 errors have been made in 200 determinations of information present or information absent, the expected value of the error rate for the population is correctly stated as "less than 0.5%."

For reporting of results to be meaningful and of practical use, it is of course necessary to report not only the determination of "information present" or "information absent" but also the statistical confidence for this individual determination (i.e., "information absent with 99% statistical confidence" for a particular subject). The bootstrapping statistical confidence for the individual result provides an estimate of the probability that a particular determination is an error. This is a more accurate, reliable, and valid predictor than the population error rate for the probability that a particular determination is an error, because it is based on the variability in this particular individual subject's data rather than on population statistics. For the determinations of "information present" and "information absent" to be statistically and practically viable, a reasonable criterion statistical confidence must be set for such that there is a high probability that each information present result and each information absent result is correct. In individual cases that lack a sufficiently high statistical confidence for either an information present or an information absent determination, the only scientifically valid outcome is to make neither determination. In such cases the outcome is in fact indeterminate and must be reported that way.

Appendix 2
Correct and Valid Application of the Bootstrapping Probability Statistic

In their authoritative seminal paper on the use of bootstrapping in psychophysiology, Wasserman and Bockenholt (1989) clearly delineated the correct use of this statistical procedure. They used Farwell and Donchin's (1988; 1991) brain fingerprinting application as an example of correct usage.

These publications and others have described the bootstrapping statistical procedure in detail. This detailed description will not be repeated here.

Very briefly, bootstrapping uses iterative sampling to estimate the distribution of a data set, to compare data of different conditions, and to estimate the probability that data meet specific criteria.

Brain fingerprinting uses bootstrapping to compute a specific probability. It computes the probability that the correlation between the probe and target responses is greater than the correlation between the probe and irrelevant responses. Recall that the target responses contain a large P300-MERMER and the irrelevant responses do not. Bootstrapping provides an answer to the question, "What is the probability that the probe responses are more similar to the target responses, which contain a P300-MERMER, than to the irrelevant responses, which do not." Correlations take into account similarities and differences in the amplitude, latency, and shape of the waveform.

The scientific significance of this is that if the proper standards have been followed this result provides evidence that the subject does or does not recognize the information contained in the probes as being known and significant in the context of the investigated situation. Information contained in the targets has this characteristic, and information contained in the irrelevants does not. The purpose of bootstrapping is to determine which of these two standards the probe response resembles, and to establish a probability that this classification is correct.

Brain fingerprinting sets a specific criterion for an "information present" determination (typically 90%), and a specific criterion for an "information absent" determination (typically 70% to 90%). If the probability is higher than the "information present" criterion that the probe responses are more similar to the target responses than to the irrelevant responses, the subject is classified as "information present." The bootstrap statistic provides the probability that this classification is correct, in light of the distribution of the single trials in the subject's data, e.g., "Determination: Information Present; Statistical Confidence: 99.9%."

Bootstrapping also computes a meaningful probability for "information absent" being the correct determination. The probability that the probe responses are more similar to the irrelevant than to the target responses is 100% minus the probability that the opposite is true, that is, "information absent" probability equals 100% minus "information present" probability. If the probability is higher than the "information absent" criterion that the probe responses are more similar to the irrelevant responses than to the target responses, the subject is classified as "information absent," e.g., "Determination: Information Absent; Statistical Confidence: 99.8%."

If the probability computed by bootstrapping does not reach either the information present or the information absent criterion, no reasonable and defensible determination can be made. The subject is classified as "indeterminate."

For use in the criminal justice, national security, or any real-world application, it is necessary to establish reasonable probability criteria for both information present and information absent determinations, and compute and report a meaningful and reasonable statistical confidence for each determination.

Some publications applying the three-stimulus paradigm used in brain fingerprinting, however, have failed to do so. They have applied the following procedure: 1) ignore the target responses; do not include them in data analysis; 2) define P300 amplitude (in one of several ways); 3) use bootstrapping to compute the probability that the probe P300 amplitude is larger than the irrelevant P300 amplitude; 4) set a specific criterion for an information-present determination, e.g., 90% probability that this determination is correct; 5) if the subject's data exceed this criterion, classify the subject as information present (sometimes referred to as "guilty"); 5) if the probability that information present is the correct determination is less than the criterion, classify the subject as information absent (or "innocent").

Bootstrapping will indeed provide a meaningful probability that the probe P300 amplitude is larger than the irrelevant P300 amplitude. Reporting a determination of "information present" with a statistical confidence of, say, 92%, that this is a correct classification of the data is both statistically and scientifically meaningful.

This procedure, however, suffers from a fatal flaw when it comes to information-absent determinations. With a 90% information-present criterion, when bootstrapping determines that there is an 89% probability that information-present is the correct determination, this procedure classifies the subject as information *absent*. The statistical procedure used has established that there is an 89% probability that the selected determination is *incorrect*; i.e., that the opposite determination is correct. The determination selected has only an 11% probability of being correct, in light of the distribution of the data as computed by bootstrapping. Obviously, an expert cannot testify in court, "We have determined that the subject is information absent, with 11% statistical confidence: There is an 11% probability that our determination is correct."

This fatal flaw cannot be eliminated by simply establishing a lower criterion for information present, for the following reason. When the true state of the subject is information absent, one expects no difference between the probe and irrelevant waveforms. Considering single trials, on average half the time the probe response will be larger and half the time the irrelevant response will be larger. If the bootstrapping procedure correctly estimates the distribution, then the expected value for the probability that the probe response is larger than the irrelevant response will be 50%. This means that, if the experiment and the statistical procedures work optimally, the average statistical confidence for a correct information-absent determination will be 50%. For subjects whose true state is information absent, half will be above and half will be below this 50% bootstrap probability. The bootstrap statistics reported in studies that use this procedure in fact approximate this situation. In other words, on average, if the procedure works optimally the statistical confidence for the correct information-absent determinations will be 50%.

This means that for information-absent subjects this procedure on average has no higher statistical confidence than a coin flip – 50%.

Changing the criterion will not solve the problem. If the criterion for an information-present determination is anything higher than 50%, then some subjects will be classified as information absent when there is a higher than 50% probability that this is the incorrect determination, and lower than 50% probability that it is the correct determination, e.g. "Determination: Information absent; Statistical confidence: 30%." If the information-present criterion is 50%, then approximately half of the subjects who are in fact information absent will be classified as information present. If the information present criterion is less than 50%, then more than half of the subjects who are in fact information absent will be classified as information present.

Even establishing an indeterminate category will not solve the problem when the bootstrap statistic is only determining the probability that the probe response is "larger" (however defined) than the irrelevant response. Say, for example, that one sets a criterion of 90% probability for an information present determination and 90% information-absent probability (equivalent to 10% information-present probability) for an information absent determination. Assume all subjects who meet neither criterion are classified as indeterminate.

If the experiment and the statistics work optimally, the determinations for subjects whose true state is information present will generally be correctly classified, because their P300 responses will be larger to the probes than to the irrelevants.

This is not the case, however, for the subjects whose true state is information absent. Since there are in fact no differences between their probe and irrelevant responses, bootstrapping if successful will return on average a 50% probability, and will only rarely deviate far from this result. Most of the subjects who are in fact information absent will be classified as indeterminate. A few of these truly information-absent subjects, randomly, may meet the 90% criterion for an information-absent determination, but this will only happen rarely and by chance. (For this to happen when only probe and irrelevant responses are considered, the bootstrapping procedure must determine there is a 90% probability that the irrelevant responses are larger than the irrelevant responses.) Given that in fact the probe and irrelevant responses are expected to be the same size, the probability that a true information-absent subject will meet the 90% criterion for an information-present determination and be misclassified as information present is equal to the probability that an information-absent subject will meet the 90% information-absent criterion and correctly be classified as information-absent. On average, the number of correct information-absent determinations will equal the number of information absent subjects incorrectly determined to be information present. Either outcome depends on a 90% probability (in one direction or the other) being computed for something for which the expected value of the probability is 50%. Here again, the procedure works no better than a coin flip for information absent subjects.

Changing the definition of a "larger" P300 will also be of no avail. Whatever the definition, for information-absent subjects the probe and irrelevant responses are expected to be the same size, and the expected value of the probability that the probe responses are larger than the irrelevant responses will be 50%. The end result in any case will be statistically no better than a coin flip.

Such a procedure, which simply compares the probe responses with the irrelevant responses, is not valid, reliable, accurate, or viable for use in the criminal justice system, national security, or any real-world application with significant consequences to the outcome of the test.

It is not necessary, however, to use correlations to compare the target, probe, and irrelevant waveforms as is done in the standard brain fingerprinting procedures. Any metric that meaningfully characterizes a P300 or P300-MERMER response can potentially be applied in a valid way. (In fact, Farwell and colleagues have experimented with a dozen other metrics in the bootstrapping procedure; so far, double-centered correlation has been the most accurate.) Any valid statistical procedure, however, must provide a high statistical confidence, not only for an information-present determination, but also for an information-absent determination when the true state of the subject is in fact information absent. To do this it is necessary to consider the target responses as well as the probe and irrelevant responses. Targets provide a standard for the presence of the tell-tale response. Irrelevants provide a standard for its absence. To obtain a valid, reliable, an practically useful result it is necessary to apply a statistic that can determine, with high probability of being correct, either that the probe responses are more similar to the target responses, which contain a large P300 or P300-MERMER, or to the irrelevant responses, which do not.