Psychophysiology, 47 (2010), 57–65. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2009.00883.x

Countermeasure mechanisms in a P300-based concealed information test

JOHN B. MEIXNER AND J. PETER ROSENFELD

Department of Psychology, Northwestern University, Evanston, Illinois, USA

Abstract

We recently introduced an accurate and countermeasure (CM)-resistant P300-based deception detection test (J.P. Rosenfeld et al., 2008). When subjects use CMs to all irrelevant items in the test, the probe P300 is increased rather than reduced, as in previous P300-based deception protocols, allowing detection of CM users. Evidence herein suggests this is partly due to an *omit* effect; the probe was the only uncountered item. Three groups were tested: a *guilty omit probe* group performed an explicit response to each irrelevant item but not to the probe, an *innocent omit irrelevant* group saw only irrelevant items and omitted a response to one item, and a *guilty no omit* group had a concealed information item as probe and performed an explicit response to each. We found a greater P300 amplitude to probes in the *guilty omit probe* condition as compared with the other two conditions, indicating a P300 enhancing effect of omitting a response to a single stimulus.

Descriptors: Deception, EEG/ERP, Countermeasures, P300, Guilty Knowledge Test, Concealed Information Test

The detection of deception has been the focus of much research in the past 20 years. Though much controversy has surrounded one deception detection protocol, the "Control Question Test" (Ben-Shakhar, 2002; National Research Council, 2003), another, the Concealed Information Test (CIT), developed by Lykken (1959, 1960), is based on scientific principles and has been well received in the scientific community. The CIT presents subjects with various stimuli, one of which is a concealed information item (such as the gun used to commit a crime). The other stimuli in the test consist of control items that are of the same class (such as other potentially deadly weapons: a knife, a bat, etc.) such that an innocent person would be unable to discriminate them from the concealed information item. If the subject's physiological response is greater for the concealed information item, then knowledge of the crime or other event is inferred.

The CIT, originally designed for autonomic nervous system testing (skin conductance, blood pressure, heart and breathing rate, etc.), specifically with the polygraph machine (Ben-Shakhar, Bar-Hillel, & Kremnitzer, 2002; Lykken, 1959, 1960), has been adapted to detect guilty knowledge using event related potentials (ERPs). Studies of this type (Allen, Iacono, & Danielson, 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1988) compare P300 ERP components elicited in subjects as they view various stimuli. P300 is known to be largest in amplitude in response to infrequently presented, personally meaningful items (Johnson, 1988; Sutton, Braren, Zubin, & John, 1965). Early P300-based CITs (Allen et al., 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1988) used test items of three types: the probe, which is the concealed information item; the irrelevant, which is of the same class as the probe but with no relevance to the crime in question; and the target, which is a randomly presented irrelevant item to which the subject must make a unique response to ensure that he/she is paying attention to the stimuli.

A number of these studies have reported high accuracy rates, that is, detecting concealed information in individual subjects at or near 90% (Allen et al., 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1988; Rosenfeld, Angell, Johnson, & Qian, 1991). However, they have also been shown to be vulnerable to countermeasures: simple, discreet mental or physical responses to irrelevant stimuli designed to convert those irrelevant stimuli into covert targets, increasing their P300 amplitudes. Thus, when a guilty subject performs a countermeasure to an irrelevant item, he creates a large P300 in response to that item. Because the primary statistical comparison of the CIT is between probe and irrelevant items, as irrelevant P300 amplitude increases, accuracy decreases. In Rosenfeld, Soskins, Bosh, and Ryan (2004) and in Mertens and Allen (2008), simple countermeasures were shown to reduce the accuracy of the P300-based CIT to 50% or less. Others have recently found poor accuracy rates even without the use of countermeasures (Lefebvre, 2006; Rosenfeld, Shue, & Singer, 2007; Mertens & Allen, 2008).

To attempt to reduce the effects of countermeasures, Rosenfeld et al. (2008) designed the *complex trial protocol* (CTP), which divides each trial into a phase containing a single probe or irrelevant stimulus followed by a second phase containing a

This research was supported by the Department of Defense Polygraph Institute Grants DODP198-P-0001 and DoDPI04-P-0002 awarded to J. Peter Rosenfeld.

Address Correspondence to: John B. Meixner, Department of Psychology, Northwestern University, 2021 Sheridan Road, Evanston, IL 60208-2700, USA. E-mail: jmeixner@northwestern.edu

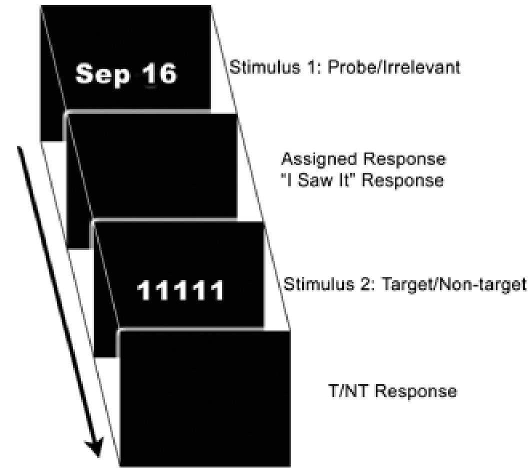

Figure 1. Structure of each trial. On each trial, subjects viewed two stimuli: one date (probe or irrelevant) and one string of numbers (target or nontarget). Using their left hand, subjects pressed an assigned response button based on the date seen, followed by the "I saw it" button for all dates. When the string of numbers appeared, subjects used their right hand to press the right mouse button if the string is all 1s (target) and the left mouse button if the string is a series of any other numbers (nontarget).

single target or nontarget stimulus (see Figure 1 for an example). This division is designed so that, for a given stimulus, the subject's attention will no longer be divided between the probe/irrelevant recognition task and the target/nontarget discrimination task because the probe/irrelevant discrimination and the target/nontarget decision tasks are separated. Rosenfeld et al. (2008) theorized that the elimination of the competing target/nontarget task should increase P300 amplitude to probe items because subjects are not dividing their attention between two different types of tasks (Donchin, Kramer, & Wickens, 1986). Rosenfeld et al. (2008) used personally meaningful information as stimuli, with items such as family surname or hometown name as the concealed information probe item. In the countermeasure block, subjects performed a distinct physical or mental countermeasure to each of four irrelevant items presented. Using this method, Rosenfeld et al. (2008) were able to detect 100% of simple guilty subjects (no CMs) as well as 92% of countermeasure users. Additionally, countermeasure users showed greater probe P300 amplitudes than simple guilty subjects, leading Rosenfeld et al. (2008) to hypothesize that the change in the protocol caused subjects executing countermeasures to focus more attention on the probe stimuli than those not executing countermeasures.

Research summarized by Donchin et al. (1986) showed that although a task competing with an oddball task results in a reduction of resources devoted to the oddball task and therefore a reduction in P300 amplitude, a second task *embedded within* an oddball task leads to an augmentation of P300 amplitude by focusing resources on the common stimulus involved in both tasks. This research led Rosenfeld et al. (2008) to hypothesize that the countermeasures used in the P300 CIT are embedded in the probe/irrelevant discrimination task in that CM execution requires increased attention to the same first stimulus discriminated in the implicit probe recognition task. In older protocols, the concurrent target/nontarget decision was a competing task, leading to reduced P300 amplitude when countermeasures were executed. In the CTP, however, the competing target/nontarget task is separated from the probe/irrelevant discrimination, allowing the embedded countermeasure task to increase the P300 response to the probe/irrelevant discrimination.

There is another non-mutually exclusive possible explanation for the results found in the countermeasure group of Rosenfeld et al. (2008). In that study, when subjects performed a countermeasure to each of the four irrelevant stimuli, the probe became additionally unique as the only stimulus not countered. Thus, when the probe appeared on the screen, subjects had to suppress the frequent action of performing a countermeasure. This response omission of an item may have made the probe even more salient, thus increasing its P300 amplitude and thereby increasing detection rates. Previous studies have found that omitting a response in a go/no-go task elicited a P300 (Roberts, Rau, Lutzenberger, & Birbaumer, 1994; Simson, Vaughan, & Ritter, 1977). However, the no-go P300s they observed were largest at central and frontal sites, whereas the classical P300 is largest at parietal midline sites (Roberts et al., 1994). It remains unknown whether the omission of countermeasure-like responses in a P300-based CIT will evoke P300 waves.

The current study seeks to extend previous work that has examined no-go P300 responses to determine the extent to which omitting a countermeasure in a P300-based CIT elicits a unique P300. Additionally, we compared the P300 amplitude elicited by omitting a countermeasure response to only one single, nonprobe item out of a set (the omit effect) to the P300 amplitude elicited by viewing personally relevant information (the probe effect). Also, we compared these P300s to the P300 amplitude of both events occurring together (probe+omit effect). It was hypothesized that the unique omission of a countermeasure to a nonprobe item in a CIT would generate a P300 similar in size to that generated by simply viewing a concealed information item in the absence of countermeasures and that the two effects combined in a single stimulus would result in a larger P300 amplitude response than that elicited by either effect alone.

Method

Participants

Thirty-eight students (average age: 19.6 years; 17 men) at Northwestern University were recruited for the study. Participants gave written informed consent to participate. Participants received either introductory psychology course credit or monetary compensation at the rate of $10 per hour for their participation. All participants had normal or corrected vision. The experiment was approved by the Northwestern Institutional Review Board.

Trial Structure

Trial structure was modeled after Rosenfeld et al. (2008). Each trial began with a 100-ms baseline period of empty black screen during which prestimulus EEG was recorded. Next, a date was presented in white text on a black background for 300 ms. Dates were presented in the form of MONTH, DAY, with the first three letters of the month used (e.g., Apr 12, Jan 23). Upon seeing the stimulus, participants were instructed to press an assigned response button using the left hand or omit an assigned response, depending on the stimulus seen (see Figure 1). Assigned responses were made using a five-button response box where participants placed each digit of the left hand on one of the buttons. Response assignments varied by group; see below. The purpose of this assigned button response task was to allow participants to mimic physical countermeasure responses that have been shown to be effective in defeating the P300-based CIT (Rosenfeld et al., 2004). Thus, these assigned responses can be thought of as overt countermeasures. Following the assigned response, participants were instructed to press the left mouse button with the right hand. Because this response indicates that the subject has seen the stimulus, regardless of whether he saw a probe or an irrelevant item, it is termed the "I saw it" button. Participants were instructed to make the assigned response with the left hand prior to pressing the "I saw it" button.

The first stimulus was followed by a randomly varying interstimulus interval of 1400 ms to 1850 ms, during which a black screen was presented to participants. Following this interval a string of six identical numbers ranging from 1 to 5 (i.e., 111111, 222222, etc.) was presented for 300 ms. Participants were instructed to press the left mouse button with the index finger of the right hand when they saw the string of ones (the target) and the right mouse button with the middle finger of the right hand when they saw a string of any other numbers (nontargets). All stimuli were shown in white font 0.7 cm high on a monitor approximately 70 cm in front of the subject.

Procedure

After signing the consent form, participants were seated in a comfortable chair and given written instructions for a practice task. The practice task was similar to the full task as described below, but included no assigned response to the first stimulus, which, for practice, was a name rather than a date. Participants were instructed to immediately press the "I saw it" button when they saw a name. Following the name, participants completed a target/nontarget recognition task as described above in the detailed trial structure. Participants practiced the task until they felt comfortable and made no errors.

Following the practice task, participants were given written instructions for the full task. Participants read these instructions and asked questions as the experimenter was applying electrodes. In the instructions, participants were told that they would be shown one of five dates and that they would have to make a unique button response using the left hand for each date that appeared. After reading the task instructions, participants were given a visual aid that showed which date was associated with each button, and participants were asked to memorize the associations. The order of the dates was presented in chronological order from left (pinkie finger) to right (thumb) in half of the participants and in reverse chronological order from left to right in the other half of the participants. Participants practiced the task until they no longer made errors on either the assigned responses or the target/nontarget responses, typically for about 5 min.

After the practice, participants completed 349 trials of the task. The task lasted approximately 30 min. The ratio of probe to irrelevant trials was 1:4, as shown in Table 1. It is noted that probe targets and probe nontargets have equal probabilities whereas irrelevant targets are much less frequent than irrelevant nontargets. This discrepancy could lead to a confound if the probability of a target following a probe being greater than that of a target following an irrelevant increased the salience of the probe item. This issue was examined in Rosenfeld et al. (2008) using a false-positive group in which the "probe" item was just another irrelevant item. If the asymmetry of conditional target probabilities caused an increase in salience of the probe, false

Table 1. *Stimulus Probabilities*

Stimulus type	Number	Probability
Probe target	33	.09
Probe nontarget	33	.09
Irrelevant target	33	.09
Irrelevant nontarget	250	.72
All probes	66	.19
All irrelevants	283	.81

Notes: Probe target ratio = .50; irrelevant target ratio = .11. A probe target trial is one in which a target follows a probe. An irrelevant target trial is one in which a target follows an irrelevant.

positive outcomes would result. Rosenfeld et al. (2008) found 0%–8% false positives, no more than in previous studies without this asymmetric probability matrix. Additionally, data from our laboratory (Rosenfeld, Tang, Meixner, Winograd, & Labkovsky, 2009) have shown that a nearly identical protocol (in which the only difference is the removal of this asymmetry) showed no difference in P300 amplitude or detection rates in comparison with the asymmetric probability protocol.

Participants were randomly assigned to one of three groups, which differed only by which buttons the participant was assigned to press for each date. The three groups were: *Guilty no omit*, *Innocent omit irrelevant*, and *Guilty omit probe*.

1. *Guilty no omit*. Participants in the guilty no omit group were shown four irrelevant dates (irrelevants) and their respective birthdate (probe). Each date was assigned a unique button on the left-hand five-button box. This group was designed to measure the effect of the probe without any unique omit response added, as all stimuli require a button press.
2. *Innocent omit irrelevant*. Participants in the innocent omit irrelevant group were shown five irrelevant dates. Four of the dates were each assigned a unique button press (pinkie through index finger; the thumb button was not used) and one of the irrelevant dates was selected to be the stimulus to which a button press was omitted. The dates for the innocent omit irrelevant group were identical to those used for the previous guilty no omit group subject, such that the probe of the previously run guilty no omit subject was the omitted item for the current innocent omit irrelevant subject. Participants were instructed to not press any button on the five-button box when they saw the date to be omitted but to still press the "I saw it" button as with any other date. Though this omitted item is not a true probe (as it is not an item of concealed information), it will hereafter be referred to as a probe item for comparison between groups. This group was designed to measure the effect of omitting a response, without any probe effect.
3. *Guilty omit probe*. Participants in the guilty omit probe group were shown four irrelevant dates (irrelevants) and the respective birthdate (probe), as in the guilty no omit group. Each irrelevant date was assigned a unique button on the left-hand five-button box (pinkie through index finger; the thumb button was not used) and the probe date was the omitted item. Participants were instructed to not press any button on the five-button box when they saw the birthdate but still to press the "I saw it" button as with any other date. This group was designed to measure the combined probe and omit effects.

Data Acquisition

Electroencephalogram (EEG) was recorded using Ag/AgCl electrodes attached to midline sites Fz, Cz, and Pz. Scalp electrodes were referenced to linked mastoids. Electrode impedances were held below 10 kΩ. Electrooculogram (EOG) was recorded differentially via Ag/AgCl electrodes placed above and below the left eye. EOG electrodes were placed diagonally to allow for the recording of both vertical and horizontal eye movements as well as eyeblinks. Artifact rejection criteria varied based on each participant's artifact amplitudes, always less than 50 μV. Trials for which this threshold was exceeded were removed from both the ERP and reaction time analyses. Data from 2 participants with fewer than 25 nonartifacted trials per stimulus were removed from the final analysis. The forehead was connected to the chassis of the isolated side of the amplifier system ("ground"). Signals were passed through Grass P511K amplifiers with a 30-Hz low-pass filter setting and high-pass filters set (3 db) at 0.3 Hz. Amplifier output was passed through a 16-bit A/D converter sampling at 500 Hz. After initial recording, single sweeps and averages were digitally filtered off-line to remove higher frequencies; 3 db point = 6 Hz.

Analysis Methods

P300 amplitude, our main dependent variable, was measured using the peak–peak method as described in Soskins, Rosenfeld, and Niendam (2001). We and others have found this analysis method to be more sensitive for the detection of deception than the standard base–peak method as used in earlier studies (Meijer, Smulders, Merckelbach, & Wolf, 2007; Soskins et al., 2001). Using in-house software designed for the Matlab platform, an algorithm searched a window of 300 ms to 650 ms to find the maximally positive segment of 100 ms, with the midpoint of this segment defined as P300 latency and its average amplitude defined as the positive P300 peak. Next, the algorithm searched a window from the P300 latency to 1300 ms to find the maximally negative segment of 100 ms. The peak–peak amplitude of the P300 was defined as the difference between the positive P300 peak and the negative P300 peak. A similar method was used to analyze the positive peak preceeding P300, hereafter referred to as P2. Based on the latency observed in the grand averages, the algorithm searched between 100 ms and 330 ms to find the maximally positive segment of 100 ms and then searched from 330 ms to 500 ms to find the maximally negative 100-ms segment. The difference between these two values was defined as the P2 peak–peak amplitude.

Within-Individuals Bootstrap Analysis

ANOVAs were applied to the Behavioral and ERP variables to assess the group effects of the study. Additionally, because this study relates to the detection of deception, individual diagnostic statistics are also essential. Though in the current study we are more interested in group effects than detection rates, it is important to examine what the effects of omitting a response to an item are within individuals in terms of hit rates, as this may help us to understand what might be the most effective method of performing countermeasures to the P300-based CIT. To determine whether the P300 evoked by a given stimulus is greater than that evoked by another stimulus within an individual, the bootstrap method (Wasserman & Bockenholt, 1989) was used at the Pz site, where P300 is usually found to be largest (Fabiani, Gratton, Karis, & Donchin, 1987). The typical bootstrap test compares the probe P300 to the average P300 of all irrelevant trials to determine whether the true difference between the average probe P300 and average irrelevant P300 is greater than zero. Because the actual distributions of probe and irrelevant waves are not available, they must be bootstrapped from the existing data. To do this, a computer program draws, with replacement, a set of individual probe waveforms equal to the number of accepted probe trials and also randomly draws (with replacement) an equal number of irrelevant waveforms. The program then subtracts the mean irrelevant P300 from the mean probe P300 and then repeats the process 100 times to create a distribution of bootstrapped probe minus irrelevant averages.

Past studies (Farwell & Donchin, 1991; Rosenfeld et al., 2004, 2008) have defined a p value criterion of .1 in order to state that a probe waveform is significantly greater than an irrelevant waveform within an individual subject. Thus, 90% of the distribution must be greater than 0 at 1.29 standard deviations from the mean of the distribution, which also means that at least 90 of the 100 iterations of the process described above must yield a positive number. In reporting bootstrap values, we report the number of iterations (out of 100) in which the probe average was greater than the irrelevant average.

Results

All within-subjects ANOVA p values reported are Greenhouse–Geisser (GG) corrected if $df > 1$. Partial eta squared values (η^2) are reported where applicable. Across all subjects, an average of 57 trials per 349 trial run (23%) were removed due to artifacts.

Behavioral: Error Rates

Error rates for the assigned response to the first stimulus (probe/irrelevant) are not discussed because two of the three groups omit a response to one of the first stimulus items, making comparison between probe and irrelevant error rates impossible. Error rates for the "I saw it" response to the first stimulus were near zero for all subjects and are likewise not discussed.

Figure 2 displays error rates for the second stimulus (target/nontarget), revealing high error rates especially for irrelevant target stimuli, as found in Rosenfeld et al. (2008). A 2 × 2 × 3 mixed model analysis of variance (ANOVA; Probe vs. Irrelevant × Target vs. Nontarget × Group) revealed a main effect of target versus nontarget, $F(1,33) = 12.85$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .28$. There was also a trend toward an interaction between probe versus irrelevant and target versus nontarget, $F(1,33) = 3.304$, $p < .08$, $\varepsilon = 1.0$, $\eta^2 = .091$. There was no main effect of probe versus irrelevant, $F(1,33) = 2.62$, $p > .115$, $\varepsilon = 1.0$, $\eta^2 = .073$, and the three way interaction was not significant, $F(2,33) = 0.304$, $p > .74$, $\varepsilon = 1.0$, $\eta^2 = .091$.

Behavioral: Reaction Times

Figure 3 shows the mean "I saw it" reaction times to both probes and the average of all irrelevants (Iall) for each group. Irrelevant reaction times (RT) were collapsed over all four irrelevant stimuli, as there was no significant difference between RT values for any single irrelevant item in a one way ANOVA, $F(3,140) = 0.786$, $p > .5$. A mixed model 2 × 3 ANOVA (Probe vs. Iall × Group) revealed a highly significant main effect of probe versus Iall, $F(1,33) = 83.2$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .716$, with probe item reaction times being faster than irrelevant item reaction times. There was also a significant interaction, $F(2,33) = 8.52$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .341$. Tukey's post hoc tests revealed differences between the guilty no omit and guilty omit probe groups

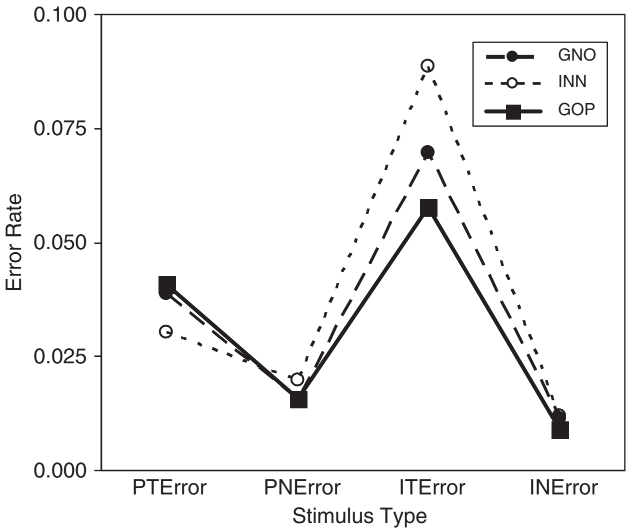

Figure 2. Group error rates to the second stimulus (target/nontarget). GNO: guilty no omit; INN: innocent omit irrelevant; GOP: guilty omit probe; PT: probe target; PN: probe nontarget; IT: irrelevant target; IN: irrelevant nontarget.

($p < .015$), with no significant differences between any other two groups. To further understand the probe "I saw it" reaction times, a one-way ANOVA was run with probe "I saw it" reaction time as the dependent variable and group as the factor. The result was highly significant, $F(2,33) = 10.59$, $p < .001$, with Tukey's post hoc tests revealing differences between guilty no omit and innocent omit irrelevant ($p < .04$) as well as between guilty no omit and guilty omit probe ($p < .001$), but no significant difference between guilty omit probe and innocent omit irrelevant ($p = .124$).

Reaction time data to the second stimulus are shown in Figure 4. A 2 × 2 × 3 mixed model ANOVA (Probe vs. Irrelevant × Target vs. Nontarget × Group) revealed significant main effects of probe versus irrelevant, $F(1,33) = 7.89$, $p < .01$, $\varepsilon = 1.0$, $\eta^2 = .193$, and target versus nontarget, $F(1,33) = 19.36$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .37$. There was also a significant interaction between probe versus irrelevant and target versus nontarget, $F(1,33) = 49.48$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .6$. There was a trend toward interactions between probe versus irrelevant and group ($p < .08$, $\eta^2 = .14$) and target versus nontarget and group ($p < .075$, $\eta^2 = .15$). The three-way interaction was not significant, $F(2,33) = 1.91$, $p = .16$, $\varepsilon = 1.0$, $\eta^2 = .1$.

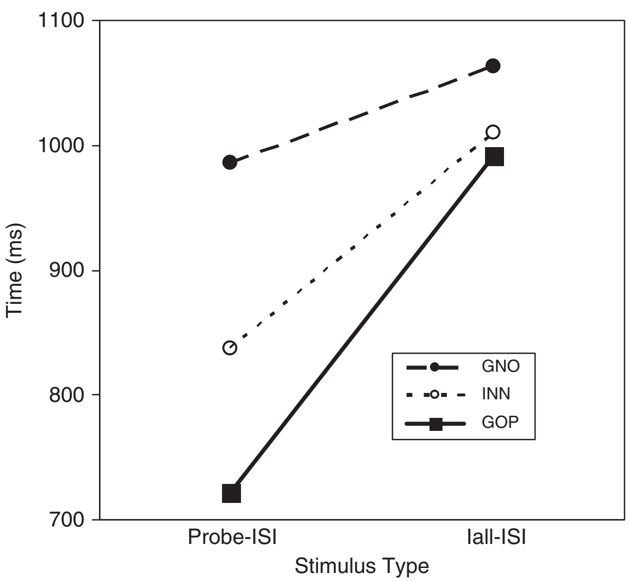

Figure 3. Group "I saw it" reaction times to the first stimulus (probe/irrelevant). GNO: guilty no omit; INN: innocent omit irrelevant; GOP: guilty omit probe. Note that on probe trials, the guilty no omit group must make an additional assigned response to probe items, explaining its large reaction time.

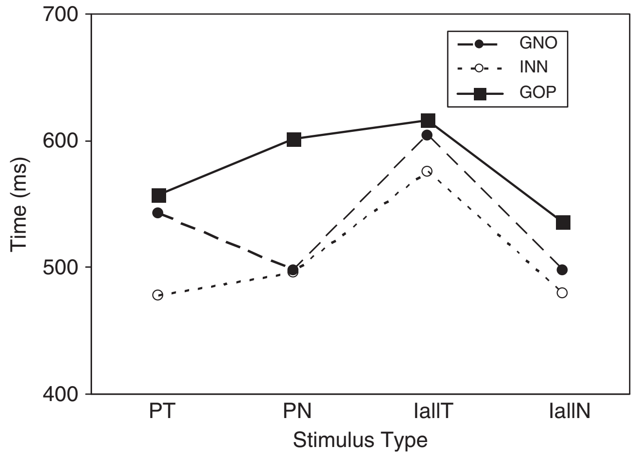

Figure 4. Group reaction times to the second stimulus. GNO: guilty no omit; INN: innocent omit irrelevant; GOP: guilty omit probe; PT: probe target; PN: probe nontarget; IT: irrelevant target; IN: irrelevant nontarget.

ERPs: Qualitative

Figure 5 shows grand average waveforms at site Pz for each group. Waveforms are shown for the probe item in each group, as well as for the average of all irrelevant items (Iall) for each group. Grand averages are restricted to the first 1200 ms of each trial, containing only the P300 response to the first stimulus (probe/irrelevant). Though the target stimuli also evoke P300s, they are of little theoretical interest in this study. As expected, large P300s are elicited by the probe items in both the guilty no omit and guilty omit probe groups as compared to the small P300s elicited by the irrelevant items. Interestingly, the guilty omit probe group P300 to the probe is much larger than that of the guilty no omit group, as supported by the quantitative group analysis below (see Figure 6). Additionally, there is a large peak–peak P300 elicited by the probe (i.e., the irrelevant item with no response as explained above) in the innocent omit irrelevant group (as compared to nonomitted irrelevant items), which will be discussed in detail below. It is noted that some subjects in the innocent omit irrelevant group had late occurring large positive peaks (between 650 and 900 ms), potentially accounting for the positive bumps seen at approximately 800 ms in the innocent omit irrelevant grand average.

ERPs: Quantitative Group Data

To examine observations about the grand averages quantitatively, a 2 × 3 × 3 (Probe vs. irrelevant × Group × Electrode) ANOVA was run (see Figure 7). Main effects were observed for each of the independent variables: probe versus irrelevant,

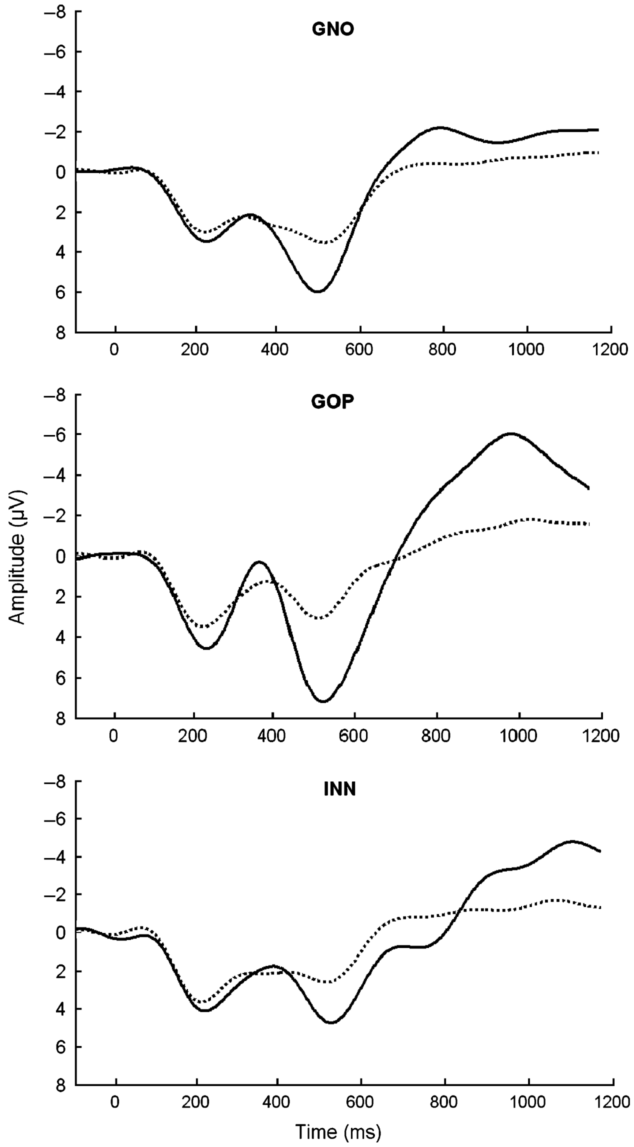

Figure 5. Grand average probe and irrelevant ERPs at Pz for each group in the study. GNO: guilty no omit; INN: innocent omit irrelevant; GOP: guilty omit probe.

$F(1,33) = 62.9$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .656$; electrode, $F(1.26, 41.5) = 7$, $p < .01$, $\varepsilon = .629$, $\eta^2 = .176$; and group, $F(2,33) = 3.2$, $p < .05$, $\eta^2 = .163$. Additionally, there was a strong trend toward an interaction between probe versus irrelevant and group, $F(2,33) = 3$, $p < .06$, $\eta^2 = .153$, and the three-way interaction between all variables was significant, $F(2.5,41.3) = 6.4$, $p < .01$, $\eta^2 = .278$. To decompose the interaction between probe versus irrelevant and group, we subtracted the Iall amplitude from the probe amplitude in each subject to compute the average probe/Iall difference for each subject averaged across all three electrode sites. A 1×3 ANOVA comparing each group's probe/Iall difference was significant, $F(2,33) = 9.4$, $p < .001$. Tukey's post hoc tests revealed significant differences between the guilty omit probe group and the innocent omit irrelevant group ($p < .001$) and between the guilty omit probe group and the guilty no omit group ($p < .01$), but there was no difference between the innocent omit irrelevant group and the guilty no omit group ($p > .9$).

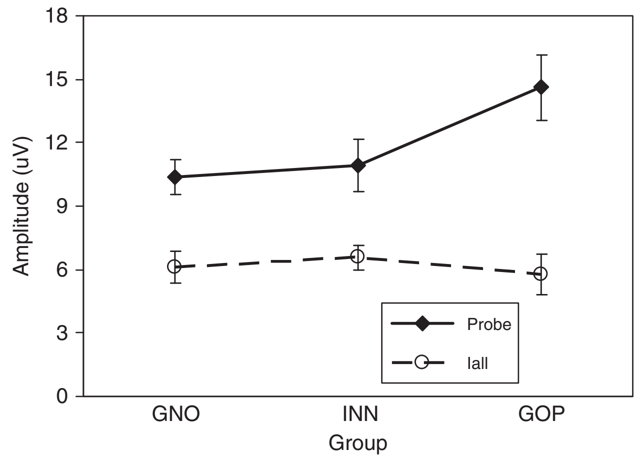

Figure 6. Average probe and irrelevant peak–peak P300 amplitudes at Pz by group. GNO: guilty no omit; INN: innocent omit irrelevant; GOP: guilty omit probe.

Because the majority of P300-based CITs are conducted using only site Pz, a 2×3 (Stimulus $\times$ Group) ANOVA was also run using data from only the Pz electrode (see Figure 6), yielding a main effect of probe versus irrelevant, $F(1,33) = 116.6$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .779$, as well as a significant interaction, $F(2,33) = 7.9$, $p < .005$, $\varepsilon = 1.0$, $\eta^2 = .325$. To decompose this interaction, we subtracted the Iall amplitude from the probe amplitude in each subject to compute the average probe/Iall difference for each subject. A 1×3 ANOVA was conducted comparing each group's probe/Iall difference, and this yielded a highly significant effect, $F(2,33) = 7.63$, $p < .005$. Tukey's post hoc tests revealed that the probe/Iall difference of the guilty omit probe group was significantly larger than that of the guilty no omit group ($p < .005$) and the innocent omit irrelevant group ($p < .01$), but there was no difference between the guilty no omit and innocent omit irrelevant groups ($p > .95$).

To analyze the P2 peak preceding P300 at site Pz (see Figure 5), a 2×3 (Probe vs. Irrelevant $\times$ Group) ANOVA was run, yielding a main effect of probe versus irrelevant, $F(1,33) = 6.6$, $p < .05$, $\varepsilon = 1.0$, $\eta^2 = .167$, but no main effect of group ($p > .05$) as well as no significant interaction. Because the effect of omitting a response does not appear to affect this peak (as evidenced by nonsignificant group effects), it will not be discussed further.

ERPs: Quantitative Individual Data

Table 2 shows detection rates within subjects for each group at the .9 confidence level as well as the number of significant iterations out of 100 in the bootstrap distribution for each subject. Bootstrap detection rates here are used to evaluate the number of individual subjects whose probe is significantly larger than the average of other stimuli. Individual subject detection rates mirror the group amplitude effects reported above, with the guilty omit probe group showing the highest detection rates (12/12) and the other two groups showing lower rates (10/12 for guilty no omit and 9/12 for innocent omit irrelevant).

Discussion

The data reported here demonstrate that the omission of a countermeasure-like discrete response to only one item in a CIT can

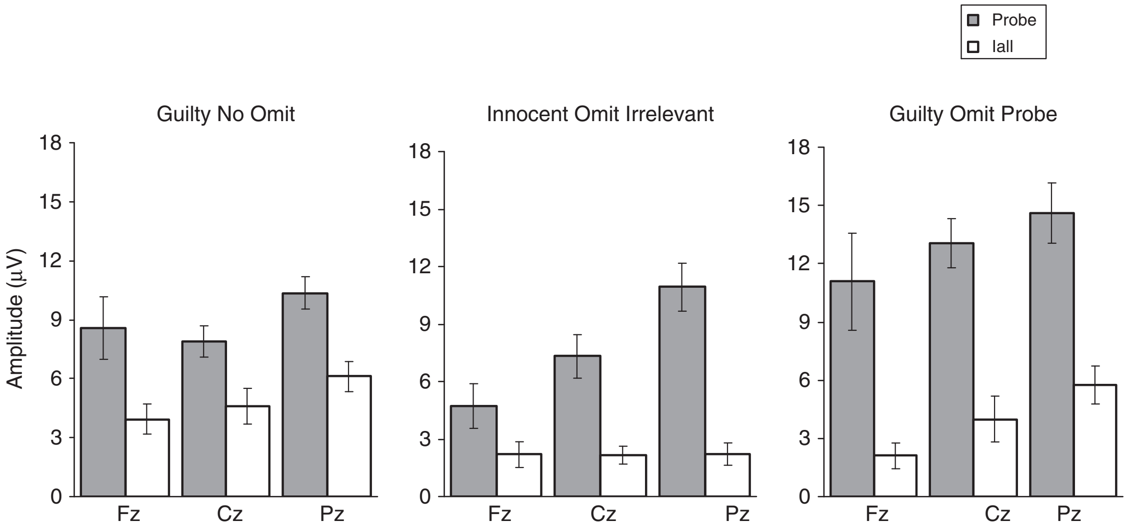

Figure 7. Average probe and irrelevant peak–peak P300 amplitudes by group at sites Fz, Cz, and Pz.

elicit a P300 response to that item. As stated previously, each of the three groups in this study was designed to measure a different P300 eliciting effect. The guilty no omit group measured the P300-eliciting effect of viewing a personally relevant stimulus (the "probe" effect), whereas the innocent omit irrelevant group measured the P300 eliciting effect of omitting a discrete button response to only a single irrelevant item (the "omit" effect). Finally, the guilty omit probe group measured both the probe and omit effects together. The peak–peak P300 elicited by the omit effect in the innocent omit irrelevant group was not statistically different than the peak–peak P300 elicited by the probe effect in the guilty no omit group (in terms of probe/omit P300 minus irrelevant P300), suggesting that the omission of a discrete response to a single stimulus may make that stimulus as salient as personally relevant information. The scalp distribution of the P300 elicited in the innocent omit irrelevant group is similar to that of the typical P300 (see Figure 7; Donchin & Coles, 1988). Moreover, the probe plus omit effect as measured by the guilty omit probe group was nearly the same size as the sum of the probe and omit effects alone as measured in the guilty no omit and innocent omit irrelevant groups (see Figure 6). This suggests that the two effects can be independently applied to the same stimulus, with each effect influencing the P300 for that stimulus, yielding a significantly larger P300 than that attained by either stimulus alone.

Table 2. *Individual Bootstrap Detection Rates of Each Subject in the Study*

Subject	GNO	INN	GOP
1	80	75	100
2	100	91	100
3	96	100	100
4	90	100	100
5	100	7	100
6	100	100	100
7	100	100	100
8	96	56	100
9	100	97	100
10	86	99	98
11	98	100	100
12	99	100	100
Total detection rate	10/12	9/12	12/12

Notes: Detection rates are based on a .9 confidence interval. Numbers indicate the number of iterations of the bootstrap process in which probe was greater than Iall. GNO: guilty no omit; INN: innocent omit irrelevant; GOP: guilty omit probe.

We suggest that the omit effect is mediated by the classical P300 oddball effect as first discovered by Sutton et al. (1965). In both the innocent omit irrelevant and guilty omit probe groups, where the omit effect is observed, subjects make a button press to each item except for one, turning this omitted item into a special target and thus eliciting a large P300. As in any classical oddball task, the subjective probability of the target (the omitted date) is much lower than that of the standard items (the irrelevant dates)—in this case 1/5 compared to 4/5. Essentially, the omit effect works similarly to how a countermeasure is intended to work, turning an irrelevant item into a target and thus amplifying its P300.

Though in this study we were primarily interested in the theoretical basis of the omit effect as it applies to CM users attempting to beat the P300-based GKT, the bootstrapped individual detection rates provide some insight into how much the omit effect can influence the probe P300 at an individual level. The guilty no omit and guilty omit probe conditions differ only by whether the subject omits a button response to the probe—the guilty omit probe group essentially combines the omit effect with the probe effect. This addition of the probe effect leads to two additional detections (see Table 2) for the guilty omit probe group as well as a large increase in the confidence level of detections: every guilty omit probe subject was detected as having a significantly larger probe P300 amplitude than irrelevant P300 amplitude at or beyond the .02 confidence level. The guilty no omit group, although still yielding 10/12 detected subjects at the .9 confidence level, does not frequently yield the same number of positive iterations of the bootstrap process that the guilty omit probe group does. In the innocent omit irrelevant group, the omit effect alone, with no influence of probe effect, was enough to make the P300 of the omitted item significantly larger than the P300s of other irrelevant items in 9/12 subjects. When this large and robust effect is combined with the probe effect as in the guilty omit probe group, a countermeasure user has little chance of generating an irrelevant P300 average large enough to compensate for the relatively enlarged probe P300 caused by the two effects together. However, because of the asymmetric probability matrix issue noted above (see Methods), as well as the fact that the groups reported in this study are not similar to the traditionally used groups in deception detection studies, the bootstrap results must be interpreted carefully. The purpose of the bootstrap statistics in the context of this study is to reveal the effect of omitting a countermeasure-like response on an individual level.

The assigned button responses in this study were designed to resemble commonly used physical countermeasures in the P300-based CIT, such as those used by Rosenfeld et al. (2004, 2008). The only major difference between the present button presses and the physical countermeasures used in the aforementioned studies is that the button presses in the current study were overt rather than covert, as true countermeasures are. This modification was made to ensure that subjects were making the correct responses within the duration of the trial as one would properly perform countermeasures. Although this does somewhat limit the ability to make conclusions based on the current data about the effects of omitting a true, covert countermeasure in a P300-based CIT, the physical and psychological mechanisms engaged by the button responses are so similar to those engaged by covert physical countermeasures that differences between the P300s elicited by the two are likely to be small.

Error rates to the second stimulus were similar to those found by Rosenfeld et al. (2008), the only published study using a similar protocol, indicating that subjects were completing the task properly and engaging similar attentional and response mechanisms to those engaged in Rosenfeld et al. (2008). The high error rate and reaction time for irrelevant target items is explained by the rare occurrence of a target following an irrelevant item as compared to the higher probability of a target following a probe. As mentioned above, the potential confounding effect of this asymmetry in probability was examined by Rosenfeld et al. (2008) with a false positive group as well as by a study from our laboratory (Rosenfeld et al., 2009). The asymmetric probability matrix was shown in both cases to have no influence on P300 amplitude to the first stimulus. Regarding the reaction times to the first stimulus (Figure 3), the only significant difference shown by the one-way ANOVAs of "I saw it" reaction time for each stimulus was between the guilty omit probe and guilty no omit groups on probe items. This and the significant interaction in the 3×2 ANOVA can be attributed to the fact that the guilty no omit group had to make an additional response to probes using the five-button response box whereas the guilty omit probe and innocent omit irrelevant groups omitted a five-button response to the probes, allowing them to press the "I saw it" button more quickly. It would be of interest to examine reaction times of each group to the first stimulus using the five-button response box, but comparisons between groups are impossible because two of the groups (guilty omit probe and innocent omit irrelevant) make no five-button response to the probe item.

The present results limit the conclusions that can be made about the effect of countermeasures in Rosenfeld et al. (2008). Using a protocol containing four irrelevant stimuli and one probe stimulus, subjects in Rosenfeld et al. (2008) were instructed to make countermeasure responses to each of the four irrelevant items while omitting a countermeasure to the probe item—highly similar to the guilty omit probe group in the current study. Subjects in the Rosenfeld et al. (2008) countermeasure group showed enhanced P300 to the countered irrelevant items, but also to the probe item, allowing most countermeasure users (23/24) to be detected by the bootstrap test. The authors hypothesized that the modifications made in the complex trial protocol caused these countermeasures to be embedded in the task of recognizing probes and irrelevants, thus increasing probe P300 (Donchin et al., 1986). Although this hypothesis may be true in part, it is difficult to know how much of the increased P300 to the probe was caused by this embedding effect and how much was caused by the omit effect when subjects omitted a countermeasure to the probe. Although this does not invalidate the claim of Rosenfeld et al. (2008) that the complex trial protocol is countermeasure resistant, it should be noted that the countermeasures used in Rosenfeld et al. (2008) did not exhaust all strategies to defeat the P300 CIT. We are currently collecting data that indicate that the complex trial protocol remains resistant to mental countermeasures to only half of the irrelevant stimuli (as opposed to countermeasures to all irrelevant stimuli as in Rosenfeld et al., 2008)—a case where no omit effect is likely to be added to the probe effect as it likely was in Rosenfeld et al. (2008). In the future it will be necessary to show that the complex trial protocol is also resistant to countermeasure strategies like these.

Evidence of the omit effect influences our understanding of the ideal countermeasures to beat a P300-based CIT. Some sophisticated perpetrators may think that the best way to counter a P300 test is to perform countermeasures to all irrelevant stimuli, thus increasing the average irrelevant P300 amplitude as much as possible so as to bring it up to the amplitude of the probe P300 created by the personal meaning of the concealed information item. Though we have previously speculated about some type of omit effect (Rosenfeld et al., 2004) and are thus conducting experiments using countermeasures to only some of the irrelevant items, the current study provides evidence that countermeasures to all will create an additional P300 enhancing effect to the noncountered probe.

REFERENCES

Allen, J., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology*, *29*, 504–522.

Ben-Shakhar, G. (2002). A critical review of the Control Questions Test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 103–126). New York: Academic Press.

Ben-Shakhar, G., Bar-Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the guilty knowledge technique in court. *Law and Human Behavior*, *26*(5).

Donchin, E., & Coles, M G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 355–372.

Donchin, E., Kramer, A., & Wickens, C. (1986). Applications of brain event related potentials to problems in engineering psychology. In M. Coles, S. Porges, & E. Donchin (Eds.), *Psychophysiology: Systems, processes and applications* (pp. 702–710). New York: Guilford.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). The definition, identification and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in Psychophysiology* (*Vol. 2*, pp. l–78). Greenwich, CT: JAI Press.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related potentials. *Psychophysiology*, *28*, 531–547.

Johnson, R. 1988. The amplitude of the P300 component of the event-related potential: Review and synthesis. In P. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology: A research annual* (*Vol. 3*, pp. 69–137). Greenwich, CT: JAI Press, Inc.

Lefebvre, C. D. (2006). *The use of event-related brain potentials (ERPs) to assess eyewitness identification accuracy*. Unpublished doctoral thesis. Halifax, NS: Dalhousie University.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*, 385–388.

Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*, 258–262.

Meijer, E. H., Smulders, F. T. Y., Merckelbach, H. L. G. J., & Wolf, A. G. (2007). The P300 is sensitive to face recognition. *International Journal of Psychophysiology*, *66*, 231–237.

Mertens, R., & Allen, J. B. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual mock crime scenarios. *Psychphysiology*, *45*, 286–298.

National Research Council (2003). *The polygraph and lie detection*. Washington DC: The National Academies Press.

Roberts, L. E., Rau, H., Lutzenberger, W., & Birbaumer, N. (1994). Mapping P300 waves onto inhibition: Go/no-go discrimination. *Electroencephalography and Clinical Neurophysiology*, *92*, 44–55.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based, control question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, *38*, 319–335.

Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *24*, 157–161.

Rosenfeld, J. P., Labkovsky, E., Lui, M. A., Winograd, M., Vandenboom, C., & Chedid, K. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate P300-based method for detection of concealed information. *Psychophysiology*, *45*, 906–919.

Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for autobiographical versus incidentally learned information. *Biological Psychology*, *74*, 396–404.

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205–219.

Rosenfeld, J. P., Tang, M., Meixner, J. B., Winograd, M., & Labkovsky, E. (2009). The effects of asymmetric versus symmetric probability of targets following probe and irrelevant stimuli in the Complex Trial Protocol (CTP) for detection of concealed information with P300. *Psychophysiology*, in press.

Simson, R., Vaughan, H. G. Jr., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual go-nogo tasks. *Electroencephalography and Clinical Neurophysiology*, *43*, 864–875.

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). The case for peak-to-peak measurement of P300 recorded at .3 Hz high pass filter settings in detection of deception. *International Journal of Psychophysiology*, *40*, 173–180.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science*, *150*, 1187–1188.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, *26*, 208–221.

(Received November 19, 2008; Accepted March 20, 2009)