Psychophysiology, 48 (2011), 149–154. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2010.01050.x

A mock terrorism application of the P300-based concealed information test

JOHN B. MEIXNER AND J. PETER ROSENFELD
Department of Psychology, Northwestern University, Evanston, Ilinois, USA

Abstract

Previous studies examining the P300-based concealed information test typically tested for mock crime or autobiographical details, but no studies have used this test in a counterterrorism scenario. Subjects in the present study covertly planned a mock terrorist attack on a major city. They were then given three separate blocks of concealed information testing, examining for knowledge of the location, method, and date of the planned terrorist attack, using the Complex Trial Protocol (Rosenfeld et al., 2008). With prior knowledge of the probe items, we detected 12/12 guilty subjects as having knowledge of the planned terrorist attack with no false positives among 12 innocent subjects. Additionally, we were able to identify 10/12 subjects and among them 20/30 crime-related details with no false positives using restricted *a priori* knowledge of the crime details, suggesting that the protocol could potentially identify future terrorist activity.

Descriptors: EEG/ERP, P300, Deception, Complex Trial Protocol, Concealed information test (CIT)

Research on the concealed information test (CIT) has typically focused on the detection of mock crime knowledge (Ben-Shakhar & Dolev, 1996; Farwell & Donchin, 1991; Lui & Rosenfeld, 2008; Lykken, 1959; Mertens & Allen, 2008; Rosenfeld et al., 1988) or personally relevant information (Lykken, 1960; Rosenfeld, Soskins, Bosh, & Ryan, 2004; Rosenfeld et al., 2008). The CIT presents subjects with various stimuli, one of which is a crime-related item (the *probe*, such as the gun used to commit a murder). Other stimuli consist of control items that are of the same class (*irrelevants*, such as other potentially deadly weapons: a knife, a bat, etc.) such that an innocent person would be unable to discriminate them from the crime-related item. If the subject's physiological response is greater for the probe item than for irrelevant items, then knowledge of the crime or other event is inferred. The CIT has since been adapted to use the P300 component as the key response (Allen, Iacono, & Danielson, 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1988). P300-based CITs have typically shown 80% to 95% accuracy in detecting guilty participants, with a 0% to 10% false positive rate (Allen et al., 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1988, 2008), though one recent study using a virtual mock crime method found guilty detection rates near 50% (Mertens & Allen, 2008). Additionally, one recent protocol has shown resistance to physical and mental countermeasures directed at irrelevant items, which have traditionally defeated the original P300-based CITs (Rosenfeld et al., 2008).

Like CITs using other physiological measures, most P300-based CIT experiments have used personally relevant information as the probe item (Rosenfeld, Shue, & Singer, 2007; Rosenfeld et al., 2008) or asked subjects to actively memorize words, testing for concealed knowledge of those words (Allen et al., 1992). Additionally, some experiments have involved the commission of mock crimes, testing for details that were central to a mock theft or espionage that subjects either actually committed (Farwell & Donchin, 1991; Lui & Rosenfeld, 2008; Rosenfeld et al., 1988) or committed in a virtual reality environment (Mertens & Allen, 2008, where subjects also had to learn the probe items verbatim prior to the commission of the virtual mock crime). However, no ERP-based CIT has yet tested subjects on crime-related information that they were not instructed to memorize for a crime that they have not yet carried out.

Previous work has shown that mere exposure to new stimuli will elicit unique ERPs for those stimuli as compared to unstudied items, even when a subject is not explicitly asked to remember the stimuli (Cycowicz & Friedman, 1999; Paller, Kutas, & McIssac, 1999). Additionally, subjects repeatedly exposed to words as they read text show differential ERPs to repeated and nonrepeated words, even when there is no instruction to memorize the repeated words (Joyce, Paller, Schwartz, & Kutas, 1999; Rugg, 1985; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). These results suggest that incidental exposure to crime-related information, without any memorization, could lead to differences between guilty and innocent participants in a CIT, though these studies did not examine individual differences, which are necessary for success in a diagnostic test like the CIT.

These results suggest potential applications of the CIT that have not been investigated, for example, as an antiterrorism tool, both as a way to identify terrorists and as a way to identify details

This research was supported by the Defense Academy for Credibility Assessment grants DODP198-P-0001, DODPI04-P-0002, and W74V8H-6-01-0001 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

PSYCHOPHYSIOLOGY SPR

of a planned terrorist attack. The current study shows the effectiveness of the P300-based CIT in identifying subjects who have planned, but not yet executed, a mock terrorist attack. Subjects spent approximately 30 min planning aspects of a terrorist attack and then underwent a CIT based on the Complex Trial Protocol as described by Rosenfeld et al. (2008). Analyses were done both given advance knowledge of the probe (to identify individuals as knowledgeable about the attack) and without advance knowledge (to identify the details of the planned attack in addition to individuals involved in the attack). Additionally, the current study uses three separate test blocks for three different categories of concealed information (location, date, and method of the planned attack) in an attempt to increase detection accuracy and reduce false positives.

Methods

Participants

Twenty-nine students (average age: 18.7 years; 14 men) at Northwestern University were recruited and gave informed consent. All subjects were right-handed. Participants received course credit for participation. All participants had normal or corrected vision.

Trial Structure

Trial structure was modeled after Rosenfeld et al. (2008). Each trial began with a 100-ms baseline period of black screen during which prestimulus electroencephalogram (EEG) was recorded. Next, depending on the block, a date, city, or method of terrorist attack was presented in white text on a black background for 300 ms (see Figure 1). Cities and methods of attack were presented as single words; months were presented as three letter abbreviations (Apr, Jan, etc). Upon seeing the stimulus, subjects pressed one of five response buttons at random, regardless of the stimulus seen. Responses were made using a five-button box where subjects placed each digit of the left hand on a separate button.

The first stimulus (*probe* or *irrelevant*) was followed by a randomly varying interstimulus interval of 1400 ms to 1850 ms, during which a black screen appeared. Following this interval, a string of six identical numbers ranging from 1 to 5 (i.e., 111111, 222222, etc.) was presented for 300 ms. Subjects 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 any other string (*nontargets*). All stimuli were shown in white font 0.7 cm high on a monitor 70 cm in front of the subject.

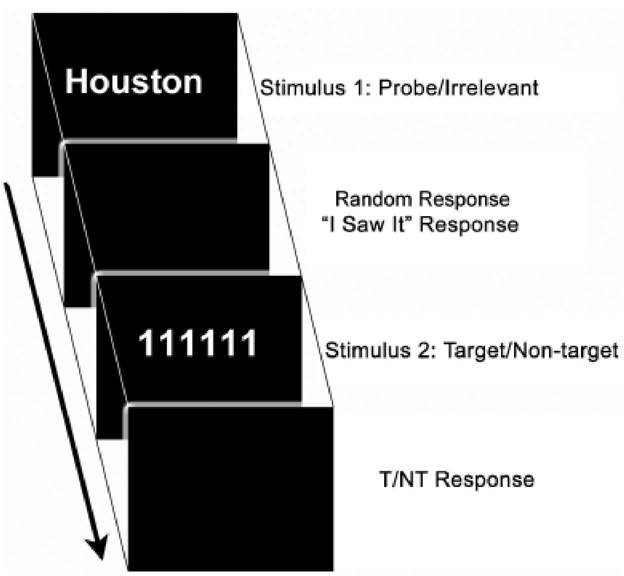

Figure 1. Trial structure.

Procedure

After signing consent, subjects were seated in a comfortable chair and given written instructions outlining the study. Among these written instructions was a form listing the cities, months, and types of terrorist attacks that would later appear in the CIT. Subjects were asked to circle any items that had personal relevance to them (e.g., the money during which month during which the subject was born, the city where the subject had lived, etc.), though they were not told they would be viewing these items later. If any of the irrelevant items had personal relevance to the subject, these items were replaced with similar items of no personal relevance.

Subjects in the guilty group ($n = 12$) were given a briefing document explaining that they were to play the role of a terrorist agent and plan a mock terrorist attack on the United States. The document detailed several different options they could choose regarding how to carry out the attack. Subjects read detailed descriptions of four types of bombs that could be used, four locations in the city of Houston that could be attacked, and four dates in July when the attack could take place. The descriptions contained pros and cons of each potential choice and instructed subjects to choose one type of bomb, one location in Houston, and one date on which to attack. After reading the briefing document, subjects were instructed to compose a letter to their superior in the terrorist organization describing the choices they had made. Note that there was no explicit formal training or instructed item memorization in this protocol. Subjects in the innocent group ($n = 12$) completed a similar task planning a vacation instead of a terrorist attack.

Subjects completed 5 min of practice, in which they performed 30 trials of a task identical to the full task as described above, except subjects viewed random first names rather than items relevant to the planned attack. The target/nontarget task in the practice was identical to that of the full task. After the practice, subjects completed three separate blocks of the task, with each block testing for a separate concealed information item. Subjects were shown potential cities (e.g., Detroit, Atlanta, etc.) where the terrorist attack could occur (with Houston to be used as the probe), potential types of terrorist attacks (with Bomb to be used as the probe) and potential months in which the attack could occur (with July to be used as the probe). Order was counterbalanced. After every 50 trials, the task was paused and subjects were asked to verbally repeat the previous item seen (to help ensure attention). Two subjects were removed from the final analysis for more than five such errors across all three blocks, and 3 subjects were removed because one of the probe items had personal relevance to them ("July" for 2 subjects, "Houston" for 1 subject), creating a confound. Each block contained 300 trials and lasted 25 min. There were five irrelevant items and one probe in each block; the ratio of probe to irrelevant trials was 1:5. Targets occurred on 10% of all trials and were equally likely to occur after either a probe or an irrelevant stimulus.

At the end of the experiment (and following all data collection), subjects were asked which city, method of attack (innocent subjects were asked about their vacation activity), and month were associated with the briefing they had read, to ensure subjects had actually read the briefing document.

Data Acquisition

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 subject's artifact amplitudes, but was always less than 50 μV. Trials for which this threshold was exceeded were removed from both the ERP and reaction time analyses. Two subjects 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 by 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 400 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 maximally negative voltage following the P300 peak. ERP analysis was only performed on the probe/irrelevant half of the trial and not on the attention enforcing target/non-target task.

Within-Individuals Bootstrap Analysis

To determine whether the P300 evoked by a given stimulus is greater than that evoked by another stimulus within an individual in each block, the bootstrap method (Wasserman & Bockenholt, 1989) was used at the Pz site, where P300 is usually largest (Fabiani, Gratton, Karis, & Donchin, 1987). 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 in each block and also draws (with replacement) an equal number of irrelevant waveforms, selected randomly from all five irrelevant items in each block. The program then subtracts the mean irrelevant P300 from the mean probe P300 and then repeats the process 1,000 times to create a distribution of bootstrapped probe minus irrelevant averages. This bootstrap test is referred to as the *Iall* test, because it compares the probe to the average of all irrelevants to determine the probability that the true difference between the average probe P300 and average irrelevant P300 is greater than zero in each block. In reporting bootstrap values, we report the number of iterations (out of 1,000) in which the probe average exceeded the irrelevant average in each block. Individual detection rates were reported based on the average number of iterations in which the probe average exceeded the irrelevant average across all three blocks. So, the bootstrap just described is conducted for each block, and each subject's three blocks are then averaged to yield the subject's bootstrap value across blocks as seen in Table 1. The maximum bootstrap value per block is 1,000, or 3,000 over the three blocks per subject. The maximum average value per subject is 3,000/3 = 1,000. For the Iall test, a .9 confidence interval cut point was used as the criterion for guilt, as in previous studies (Farwell & Donchin, 1991; Rosenfeld et al., 2004, 2008). Thus, a subject is detected as guilty if, across all three blocks, the probe average exceeds the irrelevant average (both over three blocks) on at least 900 out of 1,000 iterations of the bootstrap process.

A second, more rigorous test compared the probe P300 to the largest maximum irrelevant stimulus P300 (*Imax*). This process is identical to the Iall method, except irrelevant waveforms were drawn only from the irrelevant item that yielded the largest individual P300 amplitude. So, for example, on the city block, if "Detroit" was the irrelevant item that yielded the largest P300 amplitude for the entire block, the program would draw only from trials in which "Detroit" was the stimulus, effectively comparing the probe to the irrelevant item that generated the largest P300 amplitude. Many studies have arbitrarily used .9 confidence intervals for both Iall and Imax tests. As can be seen in the first two columns of Table 1, this may be unnecessarily stringent, as any value from .7 to .9 would yield perfect guilty–innocent

Table 1. *Individual Bootstrap Detection Rates*

Iall		Imax		Blind Imax	
Guilty	Innocent	Guilty	Innocent	Guilty	Innocent
1,000	648	985	287	985	603
1,000	610	999	416	998	602
955	598	889	476	892	649
996	611	898	430	893	605
994	150	946	17	943	689
909	475	698	284	761	547
945	600	677	365	702	536
997	555	959	250	961	569
999	586	908	217	907	565
985	690	888	382	886	706
912	390	667	129	698	650
903	644	837	215	842	702
966	**546**	**863**	**289**	**872**	**619**
12/12	0/12	12/12	0/12	10/12	0/12
AUC = 1.0		AUC = 1.0		AUC = .979	

Note: Numbers indicate the average number of iterations (across all three blocks) of the bootstrap process in which probe was greater than Iall or Imax for each of the twelve guilty and twelve innocent subjects. Blind Imax numbers indicate the average number of iterations in which the largest single item (probe or irrelevant) was greater than the second largest single item. Mean values for each column are displayed in bold above detection rates. AUC: area under the curve in the ROC analysis. Iall detection rates are based on a .9 confidence interval, Imax on a .5 interval, and Blind Imax on a .75 interval.

discrimination. Here, we use a .5 confidence interval for the Imax test (500 significant iterations or greater yields a guilty diagnosis; second two columns, Table 1) though any cut point between .5 and .65 would yield perfect guilty–innocent discrimination, as it is evident that for both guilty and especially innocent subjects, positive bootstrap iteration totals are much lower than with Iall.

This is as expected, because in an Iall bootstrap, we compare the probe item (which for an innocent subject is essentially another irrelevant item) to the average of all irrelevant items. For an innocent subject, the Iall bootstrap value should approach 500 out of 1,000 iterations because we are essentially comparing one irrelevant item to the average of the other five irrelevant items; there should be little to no difference between these values. Contrastingly, in the Imax bootstrap, we are comparing the probe item to the irrelevant item that is *by definition* the largest. Thus, for the innocent subject, we are comparing what is essentially a randomly picked irrelevant item (the probe) to the largest irrelevant item. Though these two items should be similar in size theoretically, individual variations in the ERPs for each stimulus may cause one irrelevant to be larger than another, and the Imax test by definition selects the irrelevant item that is largest, which must cause the Imax bootstrap value to be smaller as compared to Iall.

Finally, a third test (*Blind Imax*) was conducted to determine guilt or innocence if one does not know the probe *a priori*. For example, authorities may not know the city of a planned terrorist attack and might want to test a suspected terrorist conspirator for this information. To do this, one must conduct the analyses with no advance knowledge of the probe. In this test, the stimulus with the largest P300 (whether probe or irrelevant) is assumed to be the probe, and its P300 is compared with the next largest stimulus's P300, which is assumed to be the largest irrelevant P300. For this also very demanding test, it was necessary to use a lower cut point of .75 for optimal guilty–innocent discrimination. For detection of individual items in single blocks using the Blind Imax bootstrap, a .9 confidence interval was used.

It is important to note that the confidence intervals we used provide, as all such intervals should, the best guilty–innocent discrimination in the current experiment, but may not be universally ideal. The confidence intervals presented here simply indicate that there is perfect discrimination between guilty and innocent subjects when details of the attack are known ahead of time. These confidence intervals or cut points must be established with replication across many subjects. In the field, where ground truth is unknown, the ideal cut point would have to be based on well-established norms because it could not be selected *a posteriori* as was done here. It should be appreciated that there is no absolute ideal cut point in a given situation; one may reasonably use any cut point that provides an acceptable sensitivity (guilty detection rate) with an acceptable level of specificity (false positive rate).

Results

All within-subjects analysis of variance (ANOVA) p values reported are Greenhouse–Geisser (GG) corrected if $df > 1$. Partial eta squared values (η) are reported where applicable. All subjects correctly recalled all relevant details from the briefing (e.g., city, method of attack, month).

Figure 2 shows grand average waveforms at site Pz. Waveforms are shown for the probe item and the average of all irrelevant items (Iall) for both groups. Probe P300 amplitude (peak–peak) is clearly larger than Iall amplitude in the guilty group, whereas the probe and Iall amplitudes are nearly identical in the innocent group.

A 2 (Stimulus: probe vs. irrelevant) $\times$ 2 (Group: guilty vs. innocent) ANOVA was run on the peak–peak P300 amplitudes. There was a significant main effect of stimulus, $F(1,70) = 100.31$, $p < .001$, $\eta = .589$, and a main effect of group, $F(1,70) = 20.36$, $p < .001$, $\eta = .225$. The interaction was highly significant, $F(1,70) = 77.35$, $p < .001$, $\eta = .525$.

Table 1 shows detection rates within subjects for both groups, as well as the average number of significant iterations out of the maximum possible 1,000 in the bootstrap test for each subject. Detection rates are shown for each of the three tests: Iall, Imax, and Blind Imax, as described above. Perfect 12/12 detection rates with no false positives were attained when the probe item was known *a priori* in the Iall and Imax bootstrap tests. To examine the detection efficiency of each analysis method, receiver operating characteristic (ROC) analyses were conducted. The input statistic for the ROC analysis was the bootstrap statistic as displayed in Table 1. Because there is no overlap between the guilty and innocent groups for either the Iall or Imax analysis methods, the area under the curve (AUC) is 1.0, as seen in Table 1. The Blind Imax method yielded an AUC of .979.

The Blind Imax bootstrap results for each individual block are shown in Figure 3. With no *a priori* knowledge of the probe item, we were able to successfully identify 21/36 details associated with the planned terrorist attack at a .9 confidence level with no

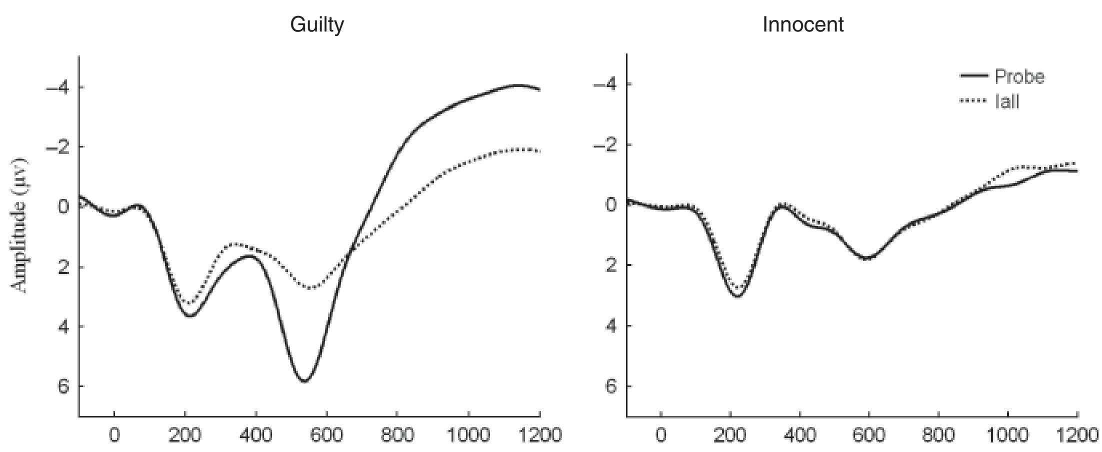

Figure 2. Grand average probe and irrelevant ERPs at Pz.

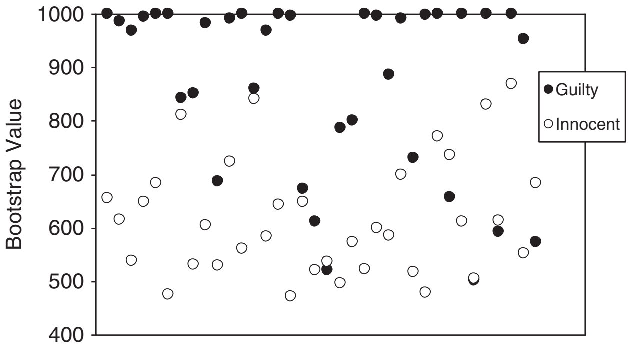

Figure 3. Bootstrap results of each individual block (12 subjects × 3 blocks = 36 guilty blocks+36 innocent blocks for 72 total blocks) using the blind Imax method. Twenty-one of 36 items associated with the terrorist attack were successfully identified at a .9 confidence level with no false positives.

false positives. ROC analysis revealed that the AUC for this analysis method was .873. Among the 10 subjects whom we successfully identified as possessing attack-related knowledge using the Blind Imax method, we were able to successfully identify 20/30 possible items as being relevant to the terrorist attack.

Discussion

The data reported here demonstrate that the Complex Trial Protocol version of the P300-based concealed information test could be highly effective in detecting an individual's knowledge of a planned terrorist attack. To the best of our knowledge, this is the first such report of a mock terrorism-based CIT. These data differ from previously reported mock crime studies (e.g., Lui & Rosenfeld, 2008; Mertens & Allen, 2008; Rosenfeld et al., 1988), because subjects in this experiment did not commit any crime—they only planned a crime that was to occur at a future date. Additionally, unlike previous studies, subjects here were not formally trained or explicitly instructed to memorize items. One mock crime study (Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003) was more similar to the current study, using a CIT to test subjects for knowledge of mock crime details that were not directly involved in the crime itself (e.g., a portrait on the wall where the crime was committed). This provided mere exposure to the probe items without explicit instructions from the experimenters to pay attention to those specific details. The results showed reduced accuracy in this more realistic type of mock crime, but this study was conducted using skin conductance response as the primary dependent measure, rather than ERPs. However, it should be noted that the depth of processing for those incidental items in Carmel et al. (2003) was likely less than that of the current study, where subjects carefully reviewed information about the planned terrorist attack.

Additionally, this study is the first to make use of multiple blocks of testing to successfully increase the sensitivity of the P300-based CIT (Rosenfeld, Shue, & Singer, 2007, used multiple blocks but achieved only 55% sensitivity using the older, non-Complex Trial Protocol version of the P300-based CIT). Though the use of multiple blocks of testing with multiple questions regarding different crime-related items is common in polygraph-based CITs (Elaad, 1990; Elaad & Ben-Shakhar, 1991; Elaad, Ginton, & Jungman, 1992), P300-based CITs have primarily focused on a single guilty knowledge item (Rosenfeld et al., 1988, 2004, 2008) or several guilty knowledge items in a single block of testing (Farwell & Donchin, 1991; Farwell & Smith, 2001; Mertens & Allen, 2008), which is problematic (Rosenfeld et al., 2004). It appears that the combination of three separate blocks of data contributed to the high individual detection accuracy in the current experiment. In the Iall and Imax bootstrap tests with 100% individual detection, many subjects had individual blocks in which the probe P300 amplitude was not significantly greater than the irrelevant amplitude such that the single block would not, by itself, yield an accurate detection. However, when the three blocks are combined, the effects of occasionally inadequate blocks are reduced, effectively increasing the signal-to-noise ratio by sampling more information from each subject.

Perhaps the most potentially useful result reported here is the moderately high rate of detection of individual blocks without specific *a priori* knowledge of the probe items (Figure 3). Allen et al. (1992) also utilized a blind Bayesian approach to the identification of learned versus unlearned lists of words. A blind approach is inherent in the Bayesian approach utilized by Allen et al., which asks about the conditional probability that a list is learned, given that it elicits a P300. This approach could have been applied to our present data set, though their approach first developed Bayesian parameters on a nonblinded basis from one sample of 20 participants and then applied those parameters to subsequent samples of subjects. The terrorist scenario used here does not readily lend itself to such a preliminary model-building phase. Additionally, subjects in the Allen et al. study were specificially instructed to memorize the lists of words that were eventually tested, which was not the case here. The lack of such explicit learning to perfection may result in ERPs with relatively lower signal-to-noise ratios pertaining to the terrorist act details that were studied only briefly. Application of the relatively more direct assumption that the actual key knowledge detail will elicit the largest P300 (measured as simple amplitude) compared in a simple bootstrap to the next largest P300 is a more direct approach that is original in the present context.

Our results suggest that one might be able to identify locations or times of terrorist attacks if the location or time is restricted to a small enough set to perform a test similar to that of the current experiment. In the field, generating such a small set may be simple for the month of the attack, as there are only 12 possible months, whereas determining the city and type of attack would be considerably more difficult because of the multitude of possibilities. In determining the city where the attack is planned to occur, one could attempt a type of partition test, where the subject is presented several potential large locations (such as the Northeast, Midwest, etc.). Using the blind analysis method demonstrated here, it should be possible to determine which of these larger areas is the planned location of attack and subsequently separate that area into smaller and smaller partitions until the location is discovered. This may not be effective, however, as a single error at any level of the process would lead the examiner astray. This type of test has not been attempted using P300 and would be an interesting future experiment.

It should be noted that 3 subjects were removed because one of the probe items had personal relevance to them. In the field, obviously, we cannot just throw out suspects for whom selected probe items have personal relevance. This is a limitation of the current study. However, there ideally would be more than three blocks of testing in the field, and extra blocks could be used to compensate for a block that is confounded by personal relevance. For example, if the probe is known to be Houston, but the suspect was born in Houston, this block could not be used, but other blocks should not be affected, still potentially allowing detection of the individual.

One does not know how well these results will translate to a field scenario, but it is likely that details that are central to a planned attack will be well rehearsed and remembered by terrorist conspirators. Although the level of encoding in the current study was comprehensive, resulting in perfect recall of the crime-relevant items when subjects were asked after the experiment, it is likely that our subjects, who spent only about 30 min learning about the attack and planning details, did not attach the same level of meaning to these items that a real terrorist would, having likely spent hours reviewing the attack plans. This increase in familiarity with the probe items could translate to larger P300s and thus greater detection efficiency in the field.

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., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, *81*, 273–281.

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: The external validity of mock crime studies. *Journal of Experimental Psychology: Applied*, *9*, 261–269.

Cycowicz, Y. M., & Friedman, D. (1999). ERP recordings during a picture fragment completion task: Effects of memory instructions. *Cognitive Brain Research*, *8*, 271–288.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, *75*, 521–529.

Elaad, E., & Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test. *International Journal of Psychophysiology*, *11*, 99–108.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, *77*, 757–767.

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. 1–78). 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.

Farwell, L. A., & Smith, S. S. (2001). Using brain MERMER testing to detect knowledge despite efforts to conceal. *Journal of Forensic Sciences*, *46*, 135–143.

Joyce, C. A., Paller, K. A., Schwartz, T. J., & Kutas, M. (1999). An electrophysiological analysis of modality-specific aspects of word repetition. *Psychophysiology*, *36*, 655–665.

Lui, M., & Rosenfeld, J. P. (2008). Detection of deception about multiple, concealed, mock crime items, based on a spatial-temporal analysis of ERP amplitude and scalp distribution. *Psychophysiology*, *45*, 721–730.

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 Psychhology*, *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. *Psychophysiology*, *45*, 286–298.

Paller, K. A., Kutas, M., & McIsaac, H. K. (1999). Monitoring conscious recollection via the electrical activity of the brain. *Psychological Science*, *6*, 107–111.

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 proble blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological Psychology*, *74*, 394–404.

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.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, *22*, 642–647.

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.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractioning the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*, 131–150.

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

(Received July 14, 2009; Accepted January 30, 2010)