Psychophysiology, 49 (2012), 1090–1100. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2012.01389.x

Combining the P300-complex trial-based Concealed Information Test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection

XIAOQING HU AND J. PETER ROSENFELD
Department of Psychology, Northwestern University, Evanston, Illinois, USA

Abstract

Despite the P300-Concealed Information Test's validity in detecting concealed memory when it is conducted immediately after the mock crime, whether the P300-CIT's detection efficiency is moderated by time delay remains unknown. Here, we conducted a mock crime study in which guilty participants were tested immediately after the mock crime or 1 month later. An innocent group was also tested. Assuming that the autobiographical Implicit Association Test (aIAT) and the P300-CIT rely on nonoverlapping mechanisms for memory detection, participants were tested using both the P300-CIT and the reaction time (RT)–based aIAT. Results suggested that the sensitivity of both tests remains even after the 1-month delay. The indicators from the RT-aIAT and P300-CIT were uncorrelated, thus combining P300-CIT and aIAT data further increased the efficiency of memory detection.

Descriptors: Event-related potentials, P300, Mock crime, Concealed Information Test, Complex trial protocol, Implicit Association Test (IAT), Autobiographical IAT, Time delay

The question of whether or not past objective memory traces can be accurately identified has received much recent attention. From an applied perspective, answering this question can shed light on uncovering the true memory status for something that an examinee is unwilling to report (e.g., lying or malingering, Hu, Hegeman, Landry, & Rosenfeld, 2012; Rosenfeld, 2011), is unable to report accurately (e.g., eye witness memory, Lefebvre, Marchand, Smith, & Connolly, 2007), or is unable to report consciously (e.g., prosopagnosia, Tranel & Damasio, 1985; see Allen, 2011). These important forensic and clinical issues have stimulated researchers to pursue various tests and measures to investigate the question.

One intensively studied protocol in memory detection is known as the Concealed Information Test (CIT; Verschuere, Ben-Shakhar, & Meijer, 2011; or the Guilty Knowledge Test, Lykken, 1959, 1960). The CIT consists of two classes of stimuli: (1) probe stimuli that pertain to the investigators' interest, such as crime-relevant information or personally significant stimuli; and (2) irrelevant stimuli that serve as comparison stimuli, such as crime-irrelevant information or personally meaningless stimuli. The probe presentation is embedded within a sequence of irrelevant stimuli, resulting in an oddball paradigm (Donchin & Coles, 1988). The assumption of the CIT is that for examinees who possess the relevant knowledge (e.g., the criminal), the probe will elicit a distinctive pattern of responses compared with those elicited by irrelevant stimuli, regardless of the examinees' overt verbal report. For people who do not have the information (e.g., innocent people who are not aware of the crime), on the other hand, the probe is just another irrelevant. Thus the responses associated with the probe should not differ from those associated with irrelevant stimuli in the innocent subjects. The CIT can utilize behavioral measures, such as reaction times (RTs, Seymour & Fraynt, 2009), autonomic nervous system (ANS) activities, such as electrodermal activity (EDA) or heart rate (Ben-Shakhar & Elaad, 2003; Gamer, Verschuere, Crombez, & Vossel, 2008), central nervous system activities, such as event-related potentials (ERPs, Allen, Iacono, & Danielson, 1992; Farwell & Donchin, 1991; Hu, Wu, & Fu, 2011; Kubo & Nittono, 2009; for a review, see Rosenfeld, 2011), or blood-oxygen level dependent (BOLD) signal responses (Gamer, Bauermann, Stoeter, & Vossel, 2007; Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Langleben et al., 2002; Spence, Farrow, Herford, Wilkinson, Zheng, & Woodruff, 2001, for a review, see Gamer, 2011b).

Despite ongoing CIT research conducted in many academic laboratories, the CIT is not used in the field (except in Japan, Osugi, 2011). Although many factors unrelated to the CIT's validity may cause this schism (such as practitioners' negative attitudes toward CIT or the difficulties in constructing a proper CIT, see Kraphol, 2011), it is acknowledged that there are real differences between laboratory-based research and field investigation. These differences include but are not limited to the following factors: (a) participants in lab studies do not have the same emotional or motivational states as do examinees in the field; and (b) the memory status of a subject investigated in the lab is usually different from that of a subject in the field: Specifically, unlike lab studies in which participants are

This research was supported by the Graduate Research Grant of Northwestern University awarded to Xiaoqing Hu.

Address correspondence to: Xiaoqing Hu, Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL, 60208. E-mail: xiaoqinghu@u.northwestern.edu

usually tested on crime-relevant information immediately after the mock crime, tests in the field may take place weeks, months, and even years after the crime. Whereas the emotion/motivation concern applies to any lab-based deception detection simulation regardless of paradigm, the memory concern applies particularly to the CIT. Thus, the present study partly focuses on exploring how the time delay between the mock crime and the test will influence the CIT's sensitivity.

Previous research, primarily conducted with ANS activities,[1] has shown that when the test is conducted with a time delay (1–2 weeks) after the mock crime, the detection ability of the CIT tends to be decreased compared with the ability of the CIT conducted immediately following the mock crime. This is attributed to the decay of memory strength as the time lag between the crime and the test is prolonged (Gamer, Kosiol, & Vossel, 2010). However, it should be noted that time delay exerts a stronger effect on peripheral items that were not directly related to the crime (e.g., a picture on the wall) than on items that were central to the crime (e.g., the stolen item or the weapon used in a murder; for central vs. peripheral items descriptions, see Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar; 2003; Gamer et al., 2010; Nahari & Ben-Shakhar, 2011; Peth, Vossel, & Gamer, 2012).

In the abovementioned ANS-CIT studies, crime-relevant information for both central and peripheral crime details were used as probes. Having multiple probes to construct a CIT is desirable but usually poses a serious challenge (Kraphol, 2011). Here we plan to use one of the most central aspects of a mock crime as a single probe, the item that was stolen, in one block of the P300-CIT test. The P300 is an ERP component that occurs from 300–800 ms after the presentation of a meaningful and/or task-relevant stimulus among a series of meaningless and/or task-irrelevant stimuli (Donchin & Coles, 1988; Johnson, 1986). The amplitude of the P300 is inversely related to task demand (Johnson, 1986). Having multiple questions for multiple probes in one block of the P300-CIT can increase participants' memory burden and task demand that will lead to smaller probe P300s and poorer detection efficiency (Rosenfeld, Shue, & Singer, 2007). Therefore, the preferred P300-CIT tests one probe from a central aspect of the crime in one block. This feature of the P300-CIT is well suited to the time-delay question: since the central aspects of the crime are less likely to be influenced by time delay, the detection efficiency of P300-CIT using such central information should not be influenced significantly by time delay. It should also be noted that when multiple items of crime-relevant information are available, it is preferred to have multiple single-probe blocks to protect innocents from false positive decisions based on chance familiarity with an arbitrary single probe.

In addition to the memory status factor, another factor that may hinder the CIT's field application is its relatively low sensitivity as reported in some studies (Carmel et al., 2003; Elaad, 1990, 2011; Elaad, Ginton, & Jungman, 1992; Kraphol, 2011; Mertens & Allen, 2008), although it always provides sufficient protection for innocents (i.e., high specificity). Researchers have aimed to increase the sensitivity of memory detection tests via two strategies. One straightforward strategy is to simultaneously record and analyze multiple physiological indexes during the CIT (e.g., EDAs, heart rates, ERPs, BOLD signals; see Ambach, Bursch, Stark, & Vaitl, 2010; Gamer & Berti, 2010; Gamer et al., 2007, 2008; Matsuda, Nittono, & Ogawa, 2011). This strategy is based on the idea that different physiological indicators may be sensitive to different cognitive processes related to the probe-irrelevant differences in the CIT (for a review, see Gamer, 2011a). Thus, a combination of measures should maximally capture the differences between probe and irrelevant, which increases the CIT's sensitivity.

Another relatively underinvestigated strategy to increase the sensitivity of memory detection involves combining data from separately administered physiologically based CITs and other types of interrogative tests. This strategy is similarly based on the idea that different tests, each of which is structured upon a different rationale, may tap into nonoverlapping psychological processes underlying memory concealment (Meijer, Smulders, Johnston, & Merckelbach, 2007; Nahari & Ben-Shakhar, 2011). For instance, Meijer et al. (2007) combined the Symptom Validity Task and the autonomic CIT to detect concealed memory. Recently, Nahari and Ben-Shakhar (2011) also used the Symptom Validity Task and the Number Guessing Test in addition to the autonomic CIT in detecting mock-crime memory. It was hypothesized that the Symptom Validity Task and Number Guessing Test, which rely on guessing, will add independent information to the CIT so as to increase the sensitivity of detecting crime-relevant or personally significant memory.

One recently developed test based on RT, the autobiographical Implicit Association Test (aIAT), which we use here together with a separately administered P300-based CIT, has been shown to be highly accurate in detecting concealed memory (Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008). The aIAT shares a highly similar rationale and structure with the original Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998), and requires participants to perform a RT-based classification task. The task consists of a series of simple (single) and double classification tasks that involve four types of sentences: (1) true sentences (e.g., *I am in front of a laptop*), (2) false sentences (e.g., *I am climbing a mountain*), (3) crime-relevant sentences (e.g., *I stole a wallet*), and (4) crime-irrelevant sentences (e.g., *I read an article*.) For detailed task structure, see Method. It is expected that for criminals, given that the crime is true, pressing the same button to crime-relevant sentences and true sentences in one block (i.e., congruent block) should be easier (less conflicting) than pressing the same button for both crime-relevant sentences and false sentences in another block (i.e., incongruent block). This difference should lead to faster RTs and fewer errors in the congruent blocks compared to RTs in the incongruent blocks for guilty subjects, and these differences would form the diagnostic basis for detecting crime-relevant memory and criminals (Sartori et al., 2008). On the other hand, given that the crime-relevant sentences are not true for innocent examinees, these subjects will show an opposite response pattern when compared with the guilty examinees.

There was a recent study of the aIAT in which ERPs were simultaneously recorded to investigate the cognitive processes (e.g., cognitive control) underlying the aIAT (Agosta, Castiello, Rigoni, Lionetti, & Sartori, 2011). Since the ERPs in this study were recorded in response to stimuli presented during the aIAT, it is likely that these neural processes are different than the P300 recorded during a separate, P300-based CIT utilizing a recognition-based oddball paradigm (Donchin & Coles, 1988). Indeed, the oddball paradigm, by definition, reduces the probability of probe stimuli, relative to irrelevant stimuli, whereas the differing stimulus classes in each block of the aIAT are equally probable. Since the

1. It should be acknowledged that researchers have also used P300 to explore this issue previously, for example, Hira, Sasaki, Matsuda, Furumitsu, and Furedy (2002); Hira (2003). However, since there were no published full-length reports, we do not include discussion of these poster session abstracts.

IAT effect is hypothesized to be built on stimulus-response compatibility manipulations (e.g., Gawronski, Deutsch, & Banse, 2011; Verschuere & De Houwer, 2011), we expect that it could add nonredundant information to the results of a P300-CIT, which relies upon stimulus recognition, and thereby improve diagnostic accuracy. The present study will be the first attempt to use the two tests together, capitalizing on their independent elements, in separate administrations for the purpose of enhancing the accuracy of detection of concealed mock crime information. Also, to the best of our knowledge, there is no previous study that investigates the aIAT's ability to detect crime-relevant memory in a delay condition. Here we will also investigate this question for the first time.

The present study thus aims to address two issues in the memory detection field: First, we will explore whether or not a time delay (1 month here) between a mock crime and a crime-relevant test will moderate the sensitivity of both the P300-based CIT and the RT-based aIAT in detecting concealed memory. Second, given that the aIAT and the P300-CIT are hypothesized to tap into different mechanisms underlying detection of concealed memory, we will investigate whether or not combining these two possibly complementary tests could lead to further increased sensitivity in identifying concealed information.

Methods

Participants

Thirty-six participants were recruited from the Northwestern University student population (23 females; age range: 18–23, *SD* = 1.76; three additional participants were excluded because of excessive movement or artifacts during the ERP sessions). Each participant was paid $15 for his/her time. The study was approved by the Northwestern Institutional Review Board.

Procedure

After providing consent forms, participants were randomly assigned to three groups: guilty participants who were tested immediately after a mock crime (immediate), guilty participants who were tested 1 month after the mock crime (delay), and innocent participants who were also tested after their arrival in the lab (innocent). Both guilty groups were asked to enact a mock crime: stealing an exam copy from a cooperating professor's (J.P.R) mailbox in the main office of the Department of Psychology. (Participants were unaware of the professor's cooperation.) For the delay group, participants were dismissed immediately after the mock crime and told that they would be called back later for the remaining tests. After approximately 1 month (range 29–39 days, *SD* = 2.39 days), these participants returned for the tests that were identical to those in the immediate and the innocent groups. Innocent participants were simply tested in the P300-CIT without committing any mock crime and were unaware of the crime-relevant information. However, before the aIAT, they were asked to help the experimenter retrieve a research article from another graduate student's room. This was to make sure that the crime-irrelevant sentences (about this act) were true for innocents and not for guilty subjects in the aIAT (see Agosta, Mega, & Sartori, 2011; implications for field use are discussed below).

For all three groups, the tests consisted of one P300-based complex trial protocol test block (CTP; Rosenfeld et al., 2008); and one autobiographical implicit association test (aIAT; Sartori et al., 2008). In the CTP, stimuli consisted of one crime probe (exam) and eight irrelevant stimuli, one of which referred to the noncriminal act performed by innocent subjects (article). The CTP was always conducted first and the aIAT second. This was because the aIAT contained only two possible alternatives (one crime-relevant detail and one crime-irrelevant detail); both were respectively presented as a probe plus one irrelevant in the P300 CTP. Thus, if a participant were to receive an aIAT prior to the CIT, they would be exposed to true crime information. If this participant then showed a large P300 to the crime-relevant information in a later P300-CIT, it could be interpreted as resulting from the previous exposure of the information during the aIAT, rather than resulting from being involved in the mock crime. In other words, previous exposure to crime-relevant information will generate a higher probability of false positives (i.e., innocents judged guilty). However, having the CTP conducted first allowed the participants to be exposed to nine stimuli, two of which would be used in the later aIAT. This may nonspecifically prime a guilty subject to be more sensitive to recognition of both the crime-relevant and crime-irrelevant sentences used in a later aIAT. However, critically, there is no way for an innocent subject to learn the correct crime-relevant information as opposed to crime-irrelevant information from the earlier CTP experience and to respond as a guilty participant during a later aIAT. This asymmetrical influence of test order on false positive occurrence required us to use the order: CTP first and aIAT second.

The P300-CTP. The P300-CTP is structured as in Hu and colleagues (2012). Each trial began with a 300-ms baseline period for the recording of prestimulus electroencephalogram (EEG). The probe or irrelevant was then presented on the center of the screen for 300 ms. Following a randomly varying interstimulus interval lasting 1,400–1,700 ms, the target/nontarget stimulus was presented also for 300 ms (for details, see below). There were 360 trials in total, consisting of a probe and eight irrelevants, each repeated 40 times, for a total of 40 probes and 320 irrelevants presented in a random order. This P300-CTP lasted for approximately 30 min.

During the task, participants first saw either a probe or one of the eight irrelevants on a given trial (see Figure 1). Participants were told to respond randomly on a five-button box by pressing one of the five buttons chosen randomly with their left hand as soon as they saw the stimulus (this response method is justified in Rosenfeld & Labkovsky, 2010). This was the stimulus acknowledgement or the "I saw it" response. They were warned that the experimenter

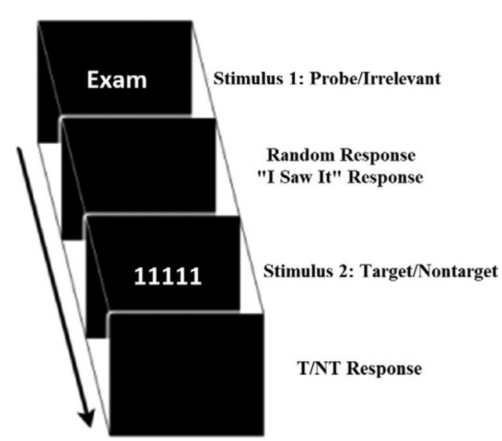

Figure 1. Task structure of the complex trial protocol.

would pause the experiment about every 20–40 trials and ask them to repeat aloud the probe or irrelevant stimulus just presented. Failure to correctly identify more than one stimulus was indicative of inattentive noncooperation, and the participant data would be dropped. However, there was no attrition for this reason. The probe/irrelevant was followed by a string of numbers (11111, 22222, 33333, 44444, and 55555). Participants were asked to make a target/nontarget decision with their right hand upon seeing the string of numbers. If the string of numbers was "11111," they were told to press the right button (target) of the response box with their middle finger, and to press the left button of the response box with their index finger if the stimulus was any other string of numbers (nontarget). The target and nontarget occurred at an equal probability following probe or following each irrelevant.

The RT-based aIAT. As noted above, after the CTP, participants in the innocent group were asked to help the experimenter obtain a research article from another graduate student's room (as opposed to the guilty participants' mock exam theft). The aIAT consisted of five blocks: (1) a simple (single) classification block in which participants press one key for true sentences and press another key for false sentences (20 trials); (2) a simple classification block (20 trials) in which participants press one key for mock-crime sentences (e.g., *I took an exam*) and press another key for non-mock-crime sentences (e.g., *I took an article*); (3) a double classification block (60 trials) in which participants press one key for either true sentences or crime-relevant sentences and press another key for either false sentences or crime-irrelevant sentences (thus this block is congruent for guilty subjects but incongruent for innocent subjects); (4) a reverse simple classification block (40 trials) in which participants reverse their button press for crime-relevant and crime-irrelevant sentences from the second block; and (5) a double classification block (60 trials) in which participants press one key for either true and crime-irrelevant sentences and another key for either false and crime-relevant sentences (thus this block is incongruent for guilty subjects but congruent for innocent subjects). The aIAT lasted for about 10–15 minutes. Given that the present study focuses on individual differences, the order of the two double classification blocks was always as described above (Gawronski et al., 2011).

Data Acquisition

EEG was recorded using Ag/AgCl electrodes attached to three midline sites: Fz, Cz, and Pz. Scalp electrodes were referenced to linked mastoids. Electrode impedance was kept below 5 kΩ. Electrooculogram (EOG) was recorded differentially via Ag/AgCl electrodes placed diagonally above and below the right eye to record vertical and horizontal eye movements as well as eye blinks. EOG voltages were called artifacts if above 50 µV, and all data from associated trials were rejected. 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 and 0.3-Hz high-pass filter (3 dB). 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; the digital filter was set up to pass frequencies from 0 to 6 Hz (3-dB point).

Analysis Methods for P300-CTP

After artifact-contaminated trials were rejected, P300 amplitude at Pz was measured using both the base-peak (b-p) and peak-peak (p-p) methods. These two measures were chosen because although they show high correlation generally, they may sometimes complement each other in identifying guilt without sacrificing specificity. These two measures were used later to examine the test's discriminative efficiency, but only p-p P300 was used in main analyses and individual diagnoses. Specifically, the algorithm searches from 300 to 900 ms for a maximal positive 100-ms segment average. The midpoint of the segment is defined as the P300 latency. For b-p measurement, the difference between this 100-ms positivity and the prestimulus 100-ms baseline is used as the P300 b-p amplitude. For p-p measurement, the algorithm continues to search from P300 latency to 1,300 ms for the maximum average 100-ms negativity. The difference between the maximal positive segment and the maximal negative segment is defined as the P300 p-p amplitude.

Analysis Methods for aIAT

For the aIAT, a D600 score was calculated as the dependent variable (Greenwald, Nosek, & Banaji, 2003; Sartori et al., 2008) as follows. First, RTs shorter than 300 ms or longer than 3,000 ms were deleted (<1%). Second, RTs (correct responses only) in congruent and incongruent blocks were averaged separately. Third, we calculated the standard deviation of the RT distributions from correct trials of congruent and incongruent blocks combined. Fourth, any incorrect responses were replaced with the mean RTs associated with that particular block plus a 600-ms penalty (see Greenwald et al., 2003). Fifth, the means of the congruent/incongruent blocks were calculated separately including incorrect responses with the error penalties. Sixth, the RT differences between the congruent and the incongruent blocks ($RT_{incongruent-congruent}$) from step five were divided by the inclusive standard deviation obtained from step three. The result of step six will be the D600 score (Greenwald et al., 2003). Usually, a positive *D* score suggests the examinee tends to associate crime-relevant sentences and true sentences (i.e., a guilty verdict) whereas a negative *D* score suggests the examinee tends to associate crime-irrelevant sentences and true sentences (i.e., an innocent verdict).

Intraindividual Bootstrap Analysis for P300-CTP

To determine whether a given participant did or did not recognize the crime-relevant information, we compared the P300 amplitude (p-p) at Pz between the average of the probe and the average of all irrelevants (Iall; Soskins, Rosenfeld, & Niendam, 2001). Since there is no actual average P300 distribution available, we used the repeated random sampling bootstrap method to draw artifact-free samples with replacement from the probe or irrelevant category (Rosenfeld, 2011). With iterations, this method allows us to obtain multiple bootstrapped averages. The procedure worked as follows: First, a computer program draws randomly, with replacement, from all accepted probe single sweeps, a set of sweeps of the same size as that of the original probe sweep set and averages them so as to obtain one individual probe average. Second, the same program draws randomly, with replacement, a number of accepted irrelevant single sweeps that is equal to the number of probes so as to obtain one bootstrapped average of irrelevants. Third, the average irrelevant P300 is subtracted from the average probe P300 to obtain a difference score. These steps are iterated 100 times to obtain a distribution of 100 such difference scores. Then, the number of iterations in which the probe averages is larger than the irrelevant averages are counted. If this number exceeds a given threshold

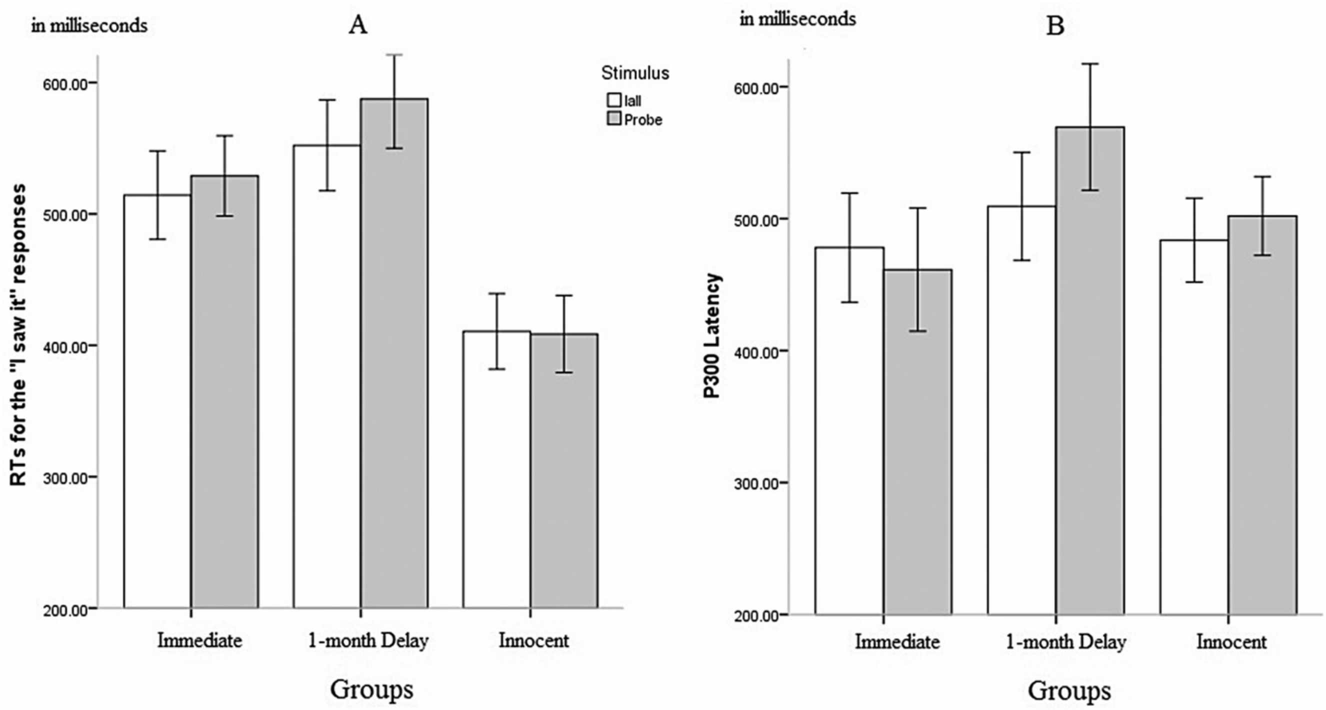

Figure 2. A: mean of the reaction times (RTs) for probe and average of all irrelevant (Iall) of the "I saw it" responses in the P300-CTP. B: P300's latency for probe and Iall in the P300-CTP. The error bar stands for one standard error.

(described below), only then is it inferred that the participants recognize the probe.

Results

All within-subject analysis of variance (ANOVA) results are reported with Greenhouse-Geisser corrected p value when $df > 1$. Partial eta squared values (η^2) are used for effect size.

Group Effects

P300-CTP. We chose three measurements as our main dependent variables in the P300-CTP: (1) RTs for the "I saw it" responses, (2) Pz-P300 latency, and (3) Pz-P300 p-p amplitude.[2] These three measures were each entered into three separate 2×3 mixed ANOVAs with stimulus type as within-subject variable (probe vs. irrelevant) and group as between-subject variable (delay vs. immediate vs. innocent).

RTs. Figure 2A shows that the RTs for probe and all irrelevants (Iall) look similar in the innocent group, but the RTs for the probe seem to be longer than RTs for the Iall in the two guilty groups, especially in the delay group. The 3×2 ANOVA on RT revealed a main effect of stimulus type $F(1,33) = 11.829$, $p < .01$, $\eta^2 = .264$, confirming that the probe RTs (mean $\pm$ *SE*, 508.306 $\pm$ 18.848 ms) were significantly longer than the Iall RTs (492.333 $\pm$ 18.714 ms). The analysis also showed a main effect of group: $F(2,33) = 6.484$, $p < .01$, $\eta^2 = .282$. Post hoc Scheffe tests revealed that the RT for the delay group was significantly longer than the RT for the innocent participants ($p < .005$), and that the RT for the immediate group was longer than the RT for the innocent participants at a marginal level of significance ($p = .065$). The stimulus Type $\times$ Group interaction was also significant, $F(2,33) = 5.381$, $p < .01$, $\eta^2 = .246$, suggesting that group status exerts different effects over the probe versus irrelevant differences as seen in Figure 2A. The follow-up paired sample tests revealed that the difference between probe and irrelevant was not significant in the innocent, $t(11) = .515$, $p > .5$, nor in the immediately tested guilty participants, $t(11) = -1.834$, $p > .09$. However, the probe had a significantly larger RT than the irrelevant RT for guilty subjects who were tested 1 month later, $t(11) = -3.287$, $p < .01$.

P300 latency. From Figure 2B, it seems that the P300 is later for probe relative to irrelevant in the delay group. However, the 3×2 ANOVA on latency did not find any main effects or interaction: $F < 2$, $p > .2$.

P300 amplitude. From Figure 3A, it is clear that the probe elicited a larger p-p P300 than the irrelevant in both the immediate and the delay groups but not in the innocent group. The 3×2 ANOVA on computed p-p amplitude from Figure 3B confirms that the main effect of stimulus type is highly significant, $F(1,33) = 38.429$, $p < .001$, $\eta^2 = .538$, with larger probe-P300 (7.066 $\pm$ .525 µV) than irrelevant-P300 (4.849 $\pm$ .409 µV). Moreover, a significant stimulus Type $\times$ Group interaction was found, $F(2,33) = 9.498$, $p < .001$, $\eta^2 = .365$. Two orthogonal contrasts were then conducted on probe-Iall P300 amplitude differences. First, the probe-Iall differences were not different between immediate and delay groups, $t(22) < 1$, $p > .9$. Second, it was found that the probe-Iall differences were significantly larger in guilty participants (collapsing immediate and delay groups) than in innocent participants, $t(34) = 4.422$, $p < .001$.

2. We also conducted the same analysis with P300 b-p amplitude: all results were the same as found with the p-p P300 amplitudes.

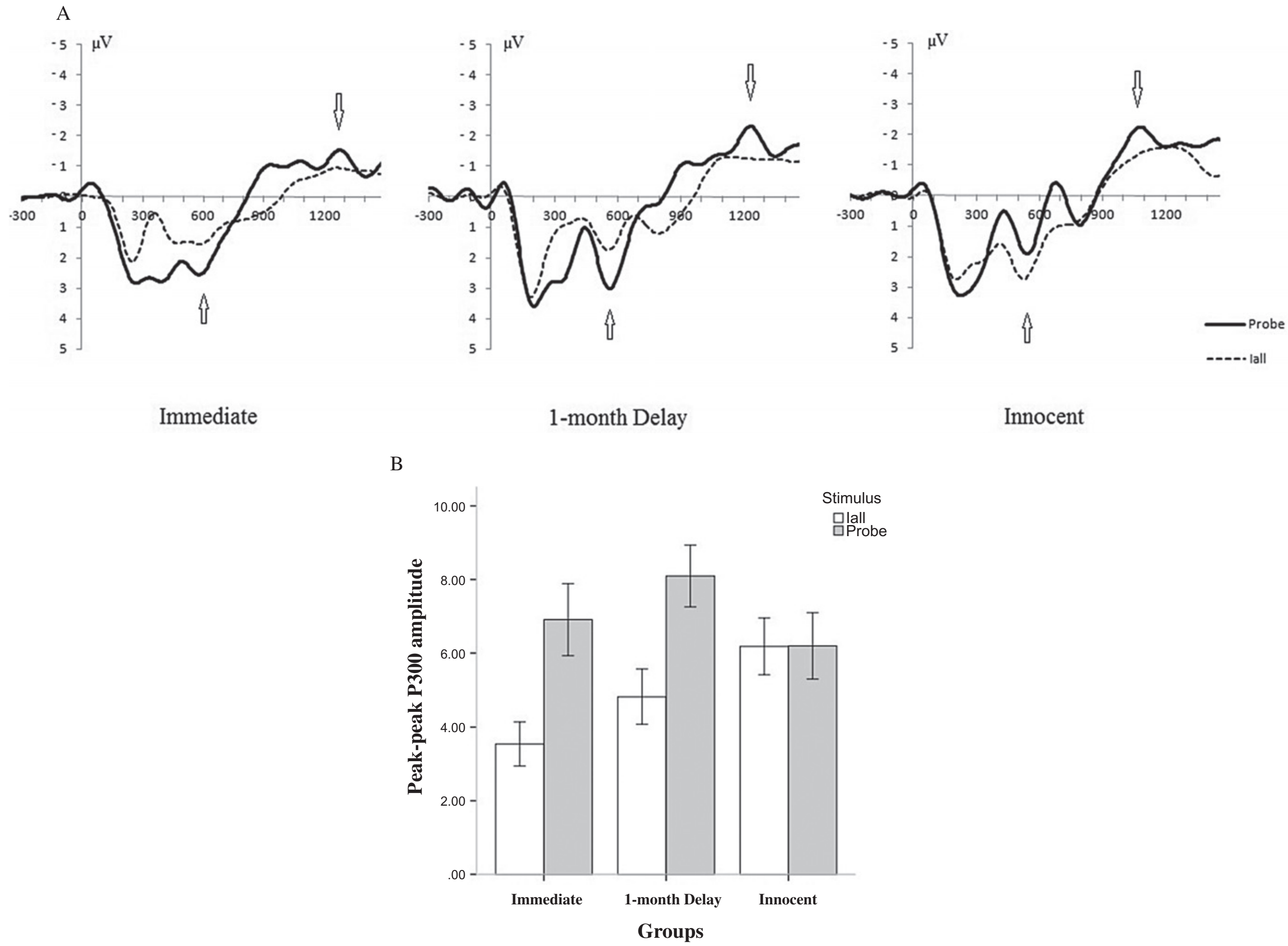

Figure 3. A: the ERP grand averages at Pz from the three groups. Positivity is downward. The up arrow indicates the P300 peak; the down arrow indicates the negativity peak. Peak-peak P300 amplitude difference is obtained as the differences between these two peaks. B: the numerical value of the average the peak-peak measured P300 at Pz from the three groups. The error bar stands for one standard error.

aIAT. Participants' RTs (based on correct trials only) across response blocks (congruent vs. incongruent blocks) and groups (immediate, delay, and innocent) are presented in Figure 4A for descriptive purposes. It is obvious that the RTs of the incongruent block (i.e., crime-false) are longer than the RTs of the congruent blocks (i.e., crime-true) for guilty subjects, but the reverse is true for innocent subjects. The *D* scores that were used for statistical analysis are presented in Figure 4B. Visual inspection shows that the *D* score was positive in the two guilty groups yet negative in the innocent group, as expected from Figure 4A. This impression was tested in a one-way ANOVA with *D* score as the dependent variable and group as the independent variable with three levels (immediate vs. 1-month delay vs. innocent). Indeed, the *D* scores changed significantly as a function of group status, $F(2,33) = 7.146$, $p < .01$. Post hoc Scheffe tests revealed that the *D* score in the innocent group was significantly lower than that in both immediate and delay groups (both $p < .01$). Moreover, the *D* scores of the immediate and delay groups were not different from each other ($p > .9$).

Correlation between electrophysiological and behavioral measures. To test the hypothesis that the P300-CTP and the RT-aIAT tap into different psychological mechanisms underlying memory detection, we conducted a correlation analysis between the dependent measures from these two tests in guilty participants. Specifically, the averaged p-p P300 amplitude difference between probe and Iall was chosen from each participant from the P300-CTP, and the participant's corresponding *D* score was chosen from the aIAT. Results showed that there is no correlation between these two indicators, $r = -.193$, $p > .37$.[3]

Detection Efficiency Based on P300-CTP and the aIAT

To determine the detection efficiency of the tests, we conducted receiver operation characteristic (ROC) analyses (Ben-Shakhar & Elaad, 2003; National Research Council, 2003). Specifically, the area under the curve (AUC) is a threshold-independent indicator of discrimination efficiency of a test considering both sensitivity (i.e., hits) and specificity (i.e., correct rejections). The AUC represents the degree of separation between the distributions of the dependent measures from guilty (here, immediate and delay groups) and inno-

3. The correlation remains nonsignificant in either the guilty-immediate group or guilty-delay group alone. Note that the Pearson correlation computation normalizes the two variables to be correlated.

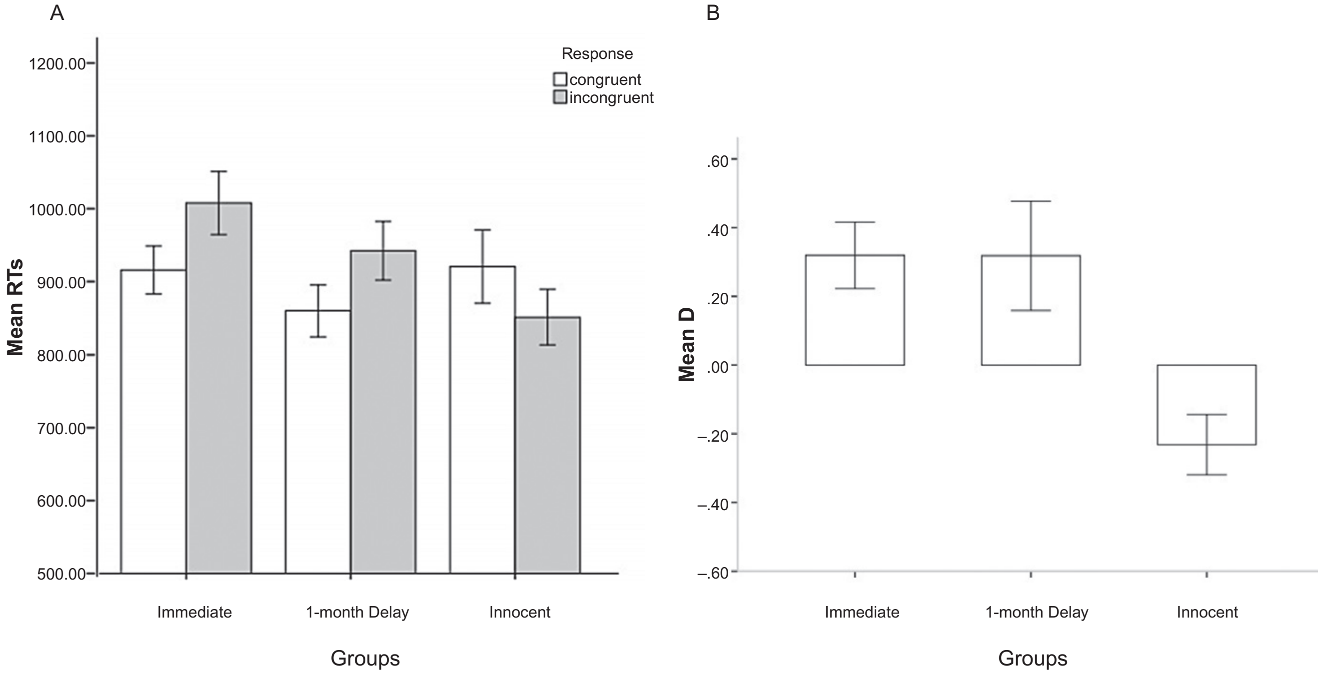

Figure 4. A: RTs (based on correct trials) from congruent and incongruent blocks of the aIAT across three groups. B: *D* score from the three groups. A positive *D* score indicates stronger associations between guilty sentences and true sentences, thus a guilty verdict. The error bar stands for one standard error.

cent participants. It varies between 0 and 1, with a chance level of 0.5 and with a perfect classification level of 1. Here, the ROC analyses were conducted based on the three indices from the P300-CTP and the *D* score from the aIAT. Specifically, the distributions of these measures from guilty participants (either immediate or delay) were compared with those from innocent participants.

CTP. The ROC analyses (see Table 1) were based on the probe-Iall differences for: (a) the RTs of the "I saw it" responses, (b) the b-p P300 amplitude, and (c) the p-p P300 amplitude. The ROC analyses showed that all three indices can accurately discriminate guilty from innocent participants above the chance (0.5) level regardless of test conditions (AUCs: 0.79–0.98, all three $ps < .001$).

aIAT. The ROC analysis based on the *D* score of the aIAT showed that the AUC values (0.79 and .90) were significantly larger than 0.5 ($ps < .001$, see Table 1), suggesting that the aIAT could effectively discriminate guilty from innocent participants across different test conditions.

Combining electrophysiological and behavioral measures. Candidate measurements from the P300-CTP (probe-Iall differences for RT for the "I saw it" responses, the P300 b-p amplitude, and the P300 p-p amplitude) and from the aIAT (*D* scores) were all transformed into *z* scores across participants. The *z* scores from the CTP were averaged as a single measure of detection efficiency of the CTP (CTP_combined). Next, *z* scores from the CTP and the *z* score from the aIAT were then averaged as a new combined measure to calculate a new ROC curve (Nahari & Ben-Shakhar, 2011). This final combined measurement showed the highest discrimination efficiency in the immediate group (from .92 to .97), although the AUC remained the same (virtual maximum) in the delay group (.99). Collapsing the immediate- and delay-guilty group, the AUC was improved from .95 to .98 (see Table 1).

Individual diagnosis. Since the b-p and p-p amplitudes are highly correlated, $r = .80$, $p < .001$, we used the bootstrap numbers based only on the more accurate p-p amplitude of P300 for individual diagnosis, as suggested by previous studies (Soskins et al., 2001). Here, we chose the bootstrap test cutoff as 85 to discriminate guilty

Table 1. *ROC Analyses Based on the P300-CTP and the RTs-aIAT*

	CTP				aIAT	Combined
Conditions	RT	b-p P300	p-p P300	CTP_combined	*D* score	All measures
Immediate	.79 (.57–.99)	.92 (.82–1.00)	.89 (.74–1.00)	.92 (.80–1.00)	.90 (.77–1.00)	.97 (.91–1.00)
Delay 1-month	.84 (.67–1.00)	.98 (.93–1.00)	.95 (.87–1.00)	.99 (.97–1.00)	.79 (.61–.97)	.99 (.97–1.00)
All guilty	.81 (.67–.96)	.95 (.89–1.00)	.92 (.83–1.00)	.95 (.89–1.00)	.84 (.71–.97)	.98 (.95–1.00)

Notes. The 95% confidence intervals of the AUCs are given in parentheses. AUC = area under the curve; RT = reaction times of the "I saw it" responses in the CTP; b-p = base-peak; p-p = peak-peak; ROC = receiver operating characteristics.

from innocent (i.e., if in 85 or more of 100 bootstrapped iterations the P300 amplitude of the probe is larger than that of Iall, the participant is judged as having recognized the probe and is inferred to be guilty; otherwise, the participant is judged as not having recognized the probe and is inferred to be innocent). This cutoff was chosen for illustration because in this study it had the perfect specificity to protect innocents (100% correct rejections), and we wanted to examine the sensitivity level at the perfect specificity level. For the *D* scores from the aIAT, we chose the cutoff of .23 to discriminate guilty from innocent. Similarly, this cutoff resulted in no false positive results (100% correct rejections). Using these cutoffs enabled us to examine whether the P300-CTP and the aIAT can complement each other to increase the sensitivity without sacrificing specificity. The numbers of participants that were correctly or incorrectly identified are presented in Table 2. It can be seen in Table 2 that the hit rates for the P300-CTP and the aIAT are largely nonoverlapping, further confirming that these two measures result from different mechanisms and are not correlated. In sum, across both immediate and delay groups, 5 out of 7 participants who were missed in the P300-CTP were correctly identified with the aIAT, whereas 10 out of 12 participants who were missed in the aIAT were correctly identified with the P300-CTP. When combining results from these two tests, if either or both of the tests showed recognition, a guilty verdict was made; if both of the tests showed nonrecognition, an innocent verdict was made. This combined diagnostic resulted in an increased sensitivity of guilt detection at 92% (22/24 in total, 11/12 in each group), and left specificity at 100%.

Discussion

The present study addressed two issues in the concealed memory detection field: (1) The effect of the delay between crime and memory detection test: We observed that 1 month after the mock crime, the P300-CTP and the aIAT seemed about equally sensitive when compared with the immediately tested guilty participants; and (2) The effect of combined P300-CIT and aIAT data on detection: We found that when the aIAT and the P300-CTP data were combined, the sensitivity of the combined tests was greater than that of either test alone.

Together with many previous P300-CIT studies (for a review, see Rosenfeld, 2011), the present results suggest that the P300-CIT is a powerful tool for crime-relevant memory detection, and has previously been shown as the only CIT to resist countermeasures (Rosenfeld, 2011). However, although it is hypothesized that memory plays a critical role in the P300 CIT, there have been no studies (see footnote 1) of the effect of the time lag between the mock crime and the test administration on the P300-CIT's detection efficiency. Moreover, given that it is unlikely that suspects in the field will be given the CIT immediately after a crime, the present results showing the CIT's detection efficiency in the time delay condition extend previous lab studies of the P300-CIT to more field-like situations. The 1-month delay here also significantly increased the time delay in comparison with those delays used previously in ANS studies (1–2 weeks, e.g., Carmel et al., 2003; Gamer et al., 2010; Lefebvre et al., 2007; Nahari & Ben-Shakhar, 2011; Peth et al., 2012). The results are promising in that we did not observe any significant decline of detection efficiency in the delay condition compared with the immediate condition. This is probably due to the fact that the item used in the present P300-CIT was the most central detail of the crime: the item that was stolen. Previous RT- and skin conductance responses (SCR)-based CITs showed that the memory strength (well-encoded central item and shallow-encoded peripheral item) and time delay work in an interactive way to influence the detection efficiency of the CIT (Gamer et al., 2010; Nahari & Ben-Shakhar, 2011; Seymour & Fraynt, 2009). Specifically, the detection efficiency remains the same across 1- or 2-week delays only when central mock crime details are used, whereas it declines when peripheral crime details are used (e.g., Gamer et al., 2010; Nahari & Ben-Shakhar, 2011; Peth et al., 2012). Thus, results from RTs-, SCR-, and P300-CITs are consistent with the idea that the memory traces of the central detail of the mock crime are less likely to decay than peripheral details and can be readily detected via the CIT.

The aIAT continues to show accurate discrimination of guilty and innocent participants. The overall reported *D* score here tends to be lower than some previously reported results (Sartori et al., 2008; but see Verschuere, Prati, & De Houwer, 2009). One possible reason for the lower *D* score here is that the aIAT was always conducted second (after the P300 test) and may have been influenced by participants' fatigue. Fatigue may have had an influence even though the present participants' RTs were within 700–1,200 ms, a range comparable with RT ranges seen in previous aIAT studies that were not preceded by a P300-CTP (e.g., Agosta, Mega, & Sartori, 2011; Hu, Rosenfeld, & Bodenhausen, in press). When we adopted the previously used zero *D* value to determine guilt or innocence (a positive *D* score indicates guilt whereas a negative *D* score indicates innocence, as in Sartori et al., 2008), the detection accuracy improved to 10/12 (from 7/12 using $D = .23$) in the immediate group, but was still only 7/12 in the delay group, and there was a cost in specificity: 4 false positives in the innocent group. Corresponding to the detection rate, the AUC associated with the aIAT was lower in the delay group (.79) than in the immediate group (0.90).[4] It is possible that the relatively low detection accuracy in the present delay and innocent groups is due to a task order effect. Indeed, in a recent study in which the aIAT (using stimuli identical to those used here) was administered immediately after the mock crime but without a preceding P300 test, the AUC was .98 (Hu et al., in press). Future studies are warranted to address this question, especially in the delay condition, without the con-

Table 2. *Number of Participants (out of 12) Classified Correctly or Incorrectly*

Conditions		aIAT Guilty	aIAT Innocent
Guilty immediate P300-CTP	Guilty	4	4
	Innocent	3	1
Guilty 1-month delay P300-CTP	Guilty	3	6
	Innocent	2	1
Innocent P300-CTP	Guilty	0	0
	Innocent	0	12

Notes. Results based on Pz peak-peak P300 amplitude bootstrap results from the P300-CTP (with 85 as cutoff) and *D* scores (with 0.23 as the cutoff) from the aIAT in the immediate, delay, and innocent groups.

4. However, this difference was not significant when we compared these two AUCs using Hanley and McNeil's (1983) method ($p > .1$). One may also notice in Table 1 that it seems as if the AUCs (for aIATs) are more affected by delay than the P300-CIT; however, the individual detection rates of the two tests in the delay group also do not differ significantly from each other based upon a Fisher's exact test ($p > .2$).

founding of order effect, as discussed below. Nonetheless, it remains encouraging that when the two sets of data from the present aIAT and P300-CTP were combined, even with the possible order effects engaged here, the results we observed were highly accurate: 92% diagnostic accuracy in guilty groups with no false positives, and AUCs greater than .95.

Although in the present study we always conducted the CIT first and the aIAT second because both tests here used the same test item (see also Nahari & Ben-Shakhar, 2011), when there are multiple crime-relevant details available, it is possible to use aIAT first for some of the details and the P300-CIT second for other details. The tests can even be administrated on two separate days to minimize fatigue or a possible task order effect. This principle similarly applies to other tests that involve two alternative choices such as the Symptom Validity Test and Number Guessing Test (e.g., in Experiment 2 of Meijer et al., 2007).

Another unexpected yet interesting delay effect was seen here with the RTs for the P300-CTP's first "I saw it" responses. Previous studies found that although this "I saw it" response does not involve any response selection, its RTs can nevertheless discriminate probe from irrelevants (Hu et al., 2012), as well as discriminating countermeasure users and nonusers (Rosenfeld & Labkovsky, 2010). Here, the RTs showed significant probe-Iall differences in the time delay group but, surprisingly, not in the immediate group. This may be because participants in the delay group had more difficulties and thus took longer to retrieve the crime-relevant memory compared with their counterparts in the immediately tested group, given the 1-month delay between the memory encoding phase (mock crime) and memory retrieval phase (memory detection test). This result also implies that although time delay may have negative influences on some indexes that are sensitive to recognition or stimulus salience (e.g., SCR), it may increase the difference between guilty and innocent on other indexes that are sensitive to stimulus evaluation or response-related processes (e.g., the RTs here).

Finally, as we expected, the P300-CIT and the aIAT complement each other well, and indeed there is no noticeable difference between the immediate and the delay groups when using combined test data. This could largely be ascribed to the putatively different mechanisms that the P300-CIT and the aIAT engage, which is consistent with the lack of correlation between these two measures. Specifically, for P300-CIT, it is the recognition of the probe that determines how well it can be discriminated from the irrelevant (Rosenfeld, 2011), while for the aIAT, it is the stimulus-response incompatibility that drives the difference between guilty and innocent participants (Gawronski et al., 2011; Verschuere & De Houwer, 2011).

Although this preliminary lab study has promise, future studies are necessary. First, it is not certain that the 85-iteration cutoff we used here for individual diagnosis based on P300 would be appropriate for the field. Future lab studies are necessary to replicate/confirm the 85 cutoff in the lab so that it might thereafter be routinely used in the field. Alternatively, in a field situation, for each test developed for each case, one could run a group of known innocent people to establish that the 85 cutoff would produce an acceptably low proportion of false positives (see Lykken, 1998).

A second issue concerns the treatment of suspects in the aIAT who are later identifiable as innocent. In the present study, we had innocents retrieve a paper as an assigned innocent act so that we could be sure that our set of crime-irrelevant sentences was true for innocent subjects. In the field, one doesn't know a priori whether or not a given suspect is innocent, so that investigators cannot assign known innocent acts as crime irrelevant sentences to suspects (as one can in the lab). However, one could assume that the alibi provided by the suspects is true for innocent but false for guilty subjects. Thus, crime-irrelevant sentences could be constructed based upon the alibi provided by every subject, whether guilty or innocent. Of course, the validity and use of such an assumption requires further rigorous investigation prior to routine field use of an aIAT.[5]

Third, whether or not the combined test can stand up to countermeasures needs to be documented. The CTP has been shown to resist countermeasures (Hu et al., 2012; Rosenfeld, 2011). The aIAT, however, has been reported to be vulnerable to faking (Verschuere et al., 2009), although Agosta, Ghirardi et al. (2011) recently reported a method for faking detection in aIAT. There is thus reason for optimism regarding the ability of combined CTP-aIAT to resist countermeasures, but this is an empirical question for future studies.

5. One may think of using the negative form of the crime-relevant sentences "*I did not steal the exam copy*" as innocent sentences. However, Agosta, Mega, and Sartori (2011) showed that using such negative sentences would lead to relatively low accuracy and reliability in identifying true memory.

References

Agosta, S., Castiello, U., Rigoni, D., Lionetti, S., & Sartori, G. (2011). The detection and the neural correlates of behavioral (prior) intentions. *Journal of Cognitive Neuroscience*, *23*, 3888–3902. doi: 10.1162/jocn_a_00039

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (2011). Detecting fakers of the autobiographical IAT. *Applied Cognitive Psychology*, *25*, 299–306. doi: 10.1002/acp.1691

Agosta, S., Mega, A., & Sartori, G. (2011). Detrimental effects of using negative sentences in the autobiographical IAT. *Acta Psychologica*, *136*, 269–275. doi: 10.1016/j.actpsy.2010.05.011

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A concealed information test with multimodal measurement. *International Journal of Psychophysiology*, *75*, 258–267. doi: 10.1016/j.ijpsycho.2009.12.007

Allen, J. 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. doi: 10.1111/j.1469-8986.1992.tb02024.x

Allen, J. J. B. (2011). Clinical applications of the Concealed Information Test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 231–252). New York, NY: Cambridge University Press.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology*, *88*, 131–151. doi: 10.1037/0021-9010.88.1.131

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. doi: 10.1037/1076-898x.9.4.261

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–427. doi: 10.1017/s0140525x00058027

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, *75*, 521–529. doi: 10.1037/0021-9010.75.5.521

Elaad, E. (2011). Validity of the Concealed Information Test in realistic contexts. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed*

Information Test (pp. 171–186). New York, NY: Cambridge University Press.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, *77*, 757–767. doi: 10.1037/0021-9010.77.5.757

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy "lie detection" with event-related brain potentials. *Psychophysiology*, *28*, 531–547. doi: 10.1111/j.1449-8986.1991.tb01990.x

Gamer, M. (2011a). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 27–45). New York, NY: Cambridge University Press.

Gamer, M. (2011b). Detecting of deception and concealed information using neuroimaging techniques. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 90–113). New York, NY: Cambridge University Press.

Gamer, M., Bauermann, T., Stoeter, P., & Vessel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Human Brain Mapping*, *28*, 1287–1301. doi: 10.1002/hbm.20343

Gamer, M., & Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology*, *47*, 355–364. doi: 10.1111/j.1469-8986.2009.00933.x

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the guilty actions test. *Biological Psychology*, *83*, 101–107, doi: 10.1016/j.biopsycho.2009.11.005

Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology & Behavior*, *95*, 333–340. doi: 10.1016/j.physbeh.2008.06.011

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex*, *13*, 830–836. doi: 10.1093/cercor/13.8.830

Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 78–123). New York, NY: Guilford Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216. doi: 10.1037/0022-3514.85.2.197

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839–843.

Hira, S. (2003). The P300-based guilty knowledge test: Does it stand the test of time? *Psychophysiology*, *40*, S10–S11.

Hira, S., Sasaki, M., Matsuda, T., Furumitsu, I., & Furedy, J. J. (2002). A year after the commission of a mock crime, the P300 amplitudes, but not reaction time, are sensitive guilty knowledge test indicators. *Psychophysiology*, *39*, S42.

Hu, X., Hegeman, D., Landry, E., & Rosenfeld, J. P. (2012). Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection. *Psychophysiology*, *49*, 85–95, doi: 10.1111/j.1469-8986.2011.01286.x

Hu, X., Rosenfeld, J. P., & Bodenhausen, G. V. (in press). Combating automatic autobiographical associations: The effect of instruction and training in strategically concealing information in the autobiographical implicit association test. *Psychological Science*.

Hu, X., Wu, H., & Fu, G. (2011). Temporal course of executive control when lying about self- and other-referential information: An ERP study. *Brain Research*, *1369*, 149–157. doi: 10.1016/j.brainres.2010.10.106

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

Kraphol, D. J. (2011). Limitations of the Concealed Information Test in criminal cases. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 151–170). New York, NY: Cambridge University Press.

Kubo, K., & Nittono, H. (2009). The role of intention to conceal in the P300-based concealed information test. *Applied Psychophysiology and Biofeedback*, *34*, 227–235. doi: 10.1007/s10484-009-9089-y

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., . . . Childress, A. R. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *NeuroImage*, *15*, 727–732. doi: 10.1006/nimg.2001.1003

Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology*, *44*, 894–904. doi: 10.1111/j.1469-8986.2007.00566.x

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*, 385–388. doi: 10.1037/h0046060

Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*, 258–262. doi: 10.1037/h0044413

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. New York, NY: Plenum Press.

Matsuda, I., Nittono, H., & Ogawa, T. (2011). Event-related potentials increase the discrimination performance of the autonomic-based concealed information test. *Psychophysiology*, *48*, 1701–1710. doi: 10.1111/j.1469-8986.2011.01266.x

Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., & Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology*, *44*, 814–822. doi: 10.1111/j.1469-8986.2007.00543.x

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

Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, *48*, 733–744. doi: 10.1111/j.1469-8986.2010.01148.x

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

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 253–275). New York, NY: Cambridge University Press.

Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, *49*, 381–390. doi: 10.1111/j.1469-8986.2011.01313.x

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 63–89). New York, NY: Cambridge University Press.

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

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

Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological Psychology*, *74*, 396–404. doi: 10.1016/j.biopsycho.2006.10.002

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science*, *19*, 772–780. doi: 10.1111/j.1467-9280.2008.02156.x

Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Applied Psychophysiology and Biofeedback*, *34*, 177–187. doi: 10.1007/s10484-009-9092-3

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioral and functional anatomical correlates of deception in humans. *NeuroReport*, *12*, 2849–2853.

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. doi: 10.1016/S0167-8760(00)00154-9

Tranel, D., & Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science*, *228*, 1453–1454. doi: 10.1126/science.4012303

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: Theory and application of the Concealed Information Test*. New York, NY: Cambridge University Press.

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: Response latency-based measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 46–62). New York, NY: Cambridge University Press.

Verschuere, B., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking in the autobiographical Implicit Association Test. *Psychological Science*, *20*, 410–413. doi: 10.1111/j.1467-9280.2009.02308.x

(Received January 24, 2012; Accepted April 11, 2012)