Research Article

Psychological Science
2014, Vol. 25(11) 1994–2005

DOI: 10.1177/0956797614547278
pss.sagepub.com

Detecting Knowledge of Incidentally Acquired, Real-World Memories Using a P300-Based Concealed-Information Test

John B. Meixner and J. Peter Rosenfeld
Northwestern University

Abstract

Autobiographical memory for events experienced during normal daily life has been studied at the group level, but no studies have yet examined the ability to detect recognition of incidentally acquired memories among individual subjects. We present the first such study here, which employed a concealed-information test in which subjects were shown words associated with activities they had experienced the previous day. Subjects wore a video-recording device for 4 hr on Day 1 and then returned to the laboratory on Day 2, where they were shown words relating to events recorded with the camera (probe items) and words of the same category but not relating to the subject's activities (irrelevant items). Electroencephalograms were recorded, and presentation of probe items was associated with a large peak in the amplitude of the P300 component. We were able to discriminate perfectly between 12 knowledgeable subjects who viewed stimuli related to their activities and 12 nonknowledgeable subjects who viewed only irrelevant items. These results have strong implications for the use of memory-detection paradigms in criminal contexts.

Keywords

autobiographical memory, memory, cognitive neuroscience, episodic memory, eyewitness memory

Received 1/1/14; Revision accepted 6/27/14

The study of memory lies at the core of psychological research. Whereas much memory research has used laboratory paradigms in which memories are acquired in a controlled, artificial setting, recent advancements in portable camera technology have enabled researchers to study retrieval of episodic memories that have been incidentally acquired throughout normal daily life. In one of the first such studies, Cabeza et al. (2004) instructed student subjects to take photos of various campus locations and subsequently measured brain activity using functional MRI while presenting to the subjects both the photos they took and similar photos taken by other subjects. Whereas presentation of both types of photo activated a similar episodic memory network (primarily medial temporal and prefrontal regions), presentation of autobiographical photos taken by the students was associated with greater activation of medial prefrontal cortex (mPFC), an area associated with self-referential processing.

More recent studies have taken a less artificial approach to studying incidentally acquired memories from normal daily life. Rather than instructing subjects to take photos, experimenters have provided subjects with a SenseCam: a small, wearable camera that records thousands of photos at regular intervals over a period of several days (Hodges, Berry, & Wood, 2011; Milton et al., 2011; St. Jacques, Conway, Lowder, & Cabeza, 2010). As did Cabeza et al. (2004), these studies have found that presentation of images from one's own daily activities is associated with increased activation in mPFC, especially ventral mPFC, both immediately after recording of the images (St. Jacques et al., 2010) and after a delay of as long as 5 months (Milton et al., 2011).

These studies have examined the psychological processes underlying autobiographical episodic memory retrieval for incidentally acquired knowledge at the group level, but they have not examined the ability to detect *individual* recognition of such incidentally acquired

Corresponding Author:
John B. Meixner, Department of Psychology, Northwestern University, 2021 Sheridan Rd., Evanston, IL 60208-2700
E-mail: j-meixner2012@nlaw.northwestern.edu

events. Detecting individual recognition of incidentally acquired knowledge has myriad potential applications but could be especially useful for the legal system, in which subjective reports of memories play a central role, especially in criminal cases. One well-known memory-detection paradigm with clear legal implications is the concealed-information test (CIT). Typically, the CIT presents subjects with various stimuli, one of which is a crime-related item (termed the probe item; such as a revolver used to commit a murder). Other stimuli consist of control items that are of the same class (termed irrelevant items; such as other types of gun: shotgun, rifle, etc.). A person without knowledge of the crime would be unable to discriminate the irrelevant items from the probe item. If the subject's physiological response is greater for the probe item than for irrelevant items, then knowledge of the crime is inferred. A number of physiological responses can be measured, such as heart rate or skin conductance (Ben-Shakhar, Bar-Hillel, & Kremnitzer, 2002); currently, one of the most effective measures is the P300 event-related-potential (ERP) component, which is large in response to meaningful, infrequently presented stimuli (Johnson, 1988; Sutton, Braren, Zubin, & John, 1965).

Most CIT research to date has focused on detecting memories through mock crime procedures (e.g., Ben-Shakhar & Dolev, 1996; Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Lui & Rosenfeld, 2008; Lykken, 1959; Meixner & Rosenfeld, 2011), in which subjects are instructed to carry out a simulated crime, which typically involves stealing a specified item out of a location, both of which are determined by the experimenters (e.g., Rosenfeld et al., 1988; Winograd & Rosenfeld, 2011), or stealing an item in a virtual environment (e.g., Mertens & Allen, 2008). Although simple to conduct, this type of method is not a realistic simulation of the detection of actual memories for several reasons. First, most mock crime studies involve a singular focus on the assigned crime and do not provide the rich array of distracting details that exist in the real world. Such distractions may decrease detection sensitivity because of reduced salience at the time of encoding. Second, the emotional salience of the crime scenario is not mimicked in a laboratory crime analogue. Third, actions committed by the subject in the lab are generally not voluntary; subjects are either told to commit a particular crime (e.g., Winograd & Rosenfeld, 2011) or given a limited number of decisions they can make regarding the commission of the crime (e.g., Rosenfeld et al., 1988).

All of these problems relate to the conditions at the moment of encoding. However, a CIT study need not test for knowledge of information encoded while the subject is in the lab itself; every individual acquires countless pieces of information each day during normal daily life that could be detected through a CIT. Detection of such everyday information may provide a more natural way to model the CIT and may reduce the impact of the three problems described in the preceding paragraph. First, because the information is acquired in a natural setting with a great amount of distracting information, the subject's focus is less likely to be directly on the items to be detected in the CIT—much as in a real crime scenario in which the perpetrator is likely to be focused on a number of things, as compared with a mock crime scenario in which the subject's focus is more singular. Second, while everyday events are not likely to capture the high level of emotion involved in the commission of a crime, they often have at least some personal importance to the subject (e.g., a heated discussion with a friend), which could trigger emotional arousal in a natural setting. Such emotional arousal in the lab context must be triggered artificially, such as through incentives to avoid detection. Third, acts conducted during normal daily life are predominantly voluntary (similar to criminal acts), not directed by experimenters.

In the present study, we drew from prior SenseCam memory studies and conducted a potentially less artificial CIT by deriving concealed-information details from the normal, voluntary daily activity of subjects. Subjects carried a small video-recording device that attached to their clothes for 4 hr, the footage from which was used to create a CIT conducted the following day. We expected that much like in mock crime scenarios, subjects taking the CIT would recognize details encountered throughout the day and thus produce a large P300 response when presented with those details.

Method

Subjects

Twenty-six students (average age: 19.9 years; 6 males, 20 females) at Northwestern University were recruited and gave informed consent following a protocol approved by the Northwestern University Institutional Review Board. All subjects were right-handed; screened by self-report for history of head injury, epilepsy, or other neurological conditions; and received $50 for participation. All reported normal or corrected-to-normal vision. Two subjects were removed from the final analysis, as described below. We selected our sample size prior to beginning the study, and the final sample of 12 subjects per group was used because prior work in this area (e.g., Meixner & Rosenfeld, 2011; Rosenfeld et al., 2008) identified this number as an appropriate sample size on the basis of formal power analyses (Rosenfeld, 2006).

Procedure

Each subject's participation in the experiment took place over 2 days. Prior to arriving in the lab, subjects were

Table 1. Probe Items Presented in the Three Trial Blocks

Subject[a]	Block 1	Block 2	Block 3
1	Class attended	Item color	Professor's name
2	Destination after lab	Computer brand used	Homework subject
3	Destination after lab	Name of friend	Sport discussed
4	Graduate school applied to	Sport discussed	Item color
5	Poster phrase	Object seen in photo	Bank Web site used
6	Exam subject	City recently visited	Recent job offer
7	Name of friend	City recently visited	Type of food eaten
8	Store shopped at	Brand of phone used	Location of lunch
9	Store shopped at	Item purchased	Location of lunch
10	Object seen in photo	Name of friend	Article recently read
11	Name of friend	Item searched for	Destination after lab
12	Class attended	Destination after lab	Item purchased

Note: Each subject in the nonknowledgeable group was yoked to a subject in the knowledgeable group, so both subjects saw the same set of stimuli.
[a]$n = 12$ because 1 subject in each group was excluded from analysis (see the text for details).

randomly assigned to either the knowledgeable group ($n = 13$) or the nonknowledgeable group ($n = 13$), though they were not told which group they were assigned to until debriefing following the CIT. The knowledgeable and nonknowledgeable groups differed only in that during the CIT, knowledgeable subjects viewed probe items that were derived from their own video footage, whereas nonknowledgeable subjects viewed probe items that were unrelated to their video footage. On Day 1, subjects met with the experimenter and were told that they would be participating in an experiment designed to measure how brain waves are influenced by decisions made in daily life. After providing consent, subjects were given written instructions outlining the study and a small video- and sound-recording device (a Muvi VVC-005; Veho, Dayton, OH) that clipped to their clothes. Subjects wore the camera for 4 hr and were told to carry out their day normally (see the Supplemental Material available online for instructions provided to subjects). Subjects then returned the camera to the lab the same day.

Once the camera was returned, an experimenter viewed the recordings to determine which information to use as probe items in the three subsequent CIT blocks. After viewing the full video, the experimenter selected three items from the total list of possible probe items and, for each probe item, developed an irrelevant set of items from the same category as the probe item. Items were selected from discrete events occurring during the video; for example, if a participant stopped at a grocery store to shop during the video, one probe item might be "grocery store" (see Table 1 for list of probe items). Irrelevant items were of approximately equal valence to the probe item (e.g., an irrelevant item for "grocery store" might be "movie theater" or "mall"). Each nonknowledgeable subject was yoked to a single knowledgeable individual subject; so, for example, Nonknowledgeable Subject 1 saw stimuli identical to those seen by Knowledgeable Subject 1. Videos were stored on a password-protected external hard drive kept in a locked room.

Subjects returned to the lab for a 2-hr testing session on Day 2. When subjects arrived, they were told that the purpose of the experiment was to test whether by using brain waves, the experimenter could determine that the subject recognized information relating to their activity while wearing the camera on Day 1. Subjects had been specifically instructed not to reveal to the experimenter any details about activity conducted while wearing the camera. While the experimenter applied the electrodes, subjects read instructions regarding the task structure (see the Supplemental Material). After finishing these instructions, subjects completed 5 min of practice of the task.

Trial structure

Trial structure was modeled after that used in Rosenfeld et al. (2008; see Fig. 1). Each trial began with a 100-ms baseline period of black screen during which prestimulus electroencephalographs (EEGs) were recorded. Next, a one- or two-word stimulus appeared. Word stimuli were related to events or information that subjects encountered when wearing the camera (e.g., items the subject may have interacted with or places the subject may have visited).

Subjects pressed a single button to indicate that they saw a stimulus appear on the screen. The first stimulus (probe or irrelevant item) was followed by a randomly varying interstimulus interval of 1,400 ms to 1,850 ms,

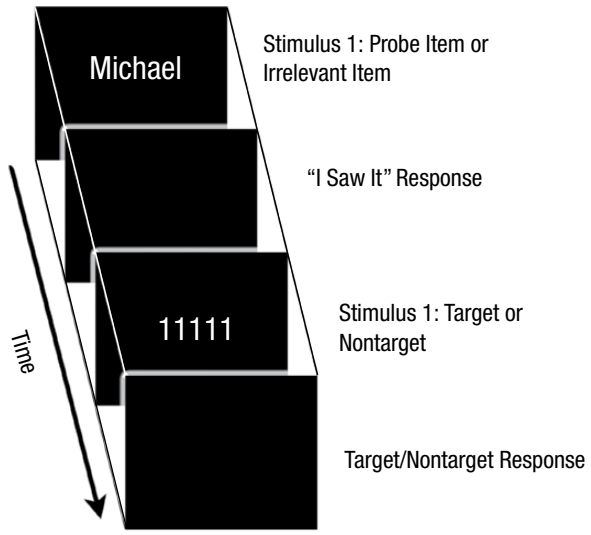

Fig. 1. Example trial sequence from Day 2 of the study. Subjects saw a word that was either related (a probe item) or unrelated (an irrelevant item) to a video they had recorded the day before. Once they indicated that they saw the stimulus appear, a string of five identical numbers ranging from 1 to 5 (e.g., "11111," "22222") was presented, and subjects had to press one button if they saw the target (a string of 1s) or another button if they saw a nontarget (any other string).

during which a black screen appeared. Following this interval, a string of five identical numbers ranging from 1 to 5 (e.g., "11111," "22222") was presented for 300 ms. Subjects were instructed to press the left mouse button with the index finger of their right hand when they saw the string of ones (the target) and the right mouse button with the middle finger of their right hand when they saw any other string (nontargets). This target/nontarget decision helped to enforce attention to the full task. All stimuli were shown in white font 0.7 cm high on a monitor 70 cm in front of the subject.

After practicing, subjects in the knowledgeable group completed three separate blocks of the task: Each block tested for a separate concealed-information item derived from the subject's video recording. Each block contained 360 trials and lasted 25 min. There were five irrelevant items and one probe item in each block, and each item was presented 60 times. Target numbers occurred on 10% of all trials and were equally likely to occur after either a probe or an irrelevant item.

Prior to each block, subjects were given a sheet of paper identifying each of the six items they would be shown in that block (one probe and five irrelevant items). Subjects were instructed to indicate whether any of the irrelevant items were in fact personally relevant to them. Whenever they identified such an item, it was replaced (10 replacements were made for knowledgeable subjects, 8 for nonknowledgeable subjects). To ensure that subjects attended to each stimulus, we occasionally asked them (on average, once every 50 trials) to report the most recently presented stimulus. Subjects with two or more errors in response to these questions would have been removed from the final analysis, but no subjects reached this threshold.

Subjects in the nonknowledgeable group completed an identical procedure to that completed by the knowledgeable group, except that probe items were not relevant to anything they did while wearing the camera but were instead the stimuli derived from a paired knowledgeable subject. The purpose of this design was to model what would occur in a real investigation, in which a knowledgeable suspect and a nonknowledgeable suspect of the same crime would be shown identical crime-related details.

Following the task, all subjects were asked whether they recognized any of the stimuli as relevant to something they had done while wearing the camera on Day 1. All knowledgeable subjects correctly recalled the three probe items. One nonknowledgeable subject informed the experimenter, following the CIT, that one of the probe items had personal relevance to her, so this subject was removed from the analysis. Thus, data from only 12 subjects in the nonknowledgeable group were analyzed.

Data acquisition

EEG data were recorded using Ag/AgCl electrodes attached to midline sites Fz, Cz, and Pz. Scalp electrodes were referenced to linked mastoids.[1] Electrode impedances were held below 10 kΩ. Electrooculogram (EOG) data were 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 eye blinks. Criteria for rejection on the basis of EOG artifacts varied according to each subject's artifact amplitudes and were always less than 50 μV. Trials for which this threshold was exceeded were removed from both the ERP and reaction-time analyses.

The forehead was connected to the chassis of the isolated side of the amplifier system ("ground"). Signals were passed through Grass (Warwick, RI) P511K amplifiers that had low-pass filters set at 30 Hz and high-pass filters set (3 db) at 0.3 Hz. Amplifier output was passed through a 16-bit analog-to-digital 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). Criteria for rejection on the basis of EEG artifacts varied according to each subject's artifact amplitudes and were always less than 100 μV. Trials for which this threshold was exceeded were removed from both the ERP and reaction-time analyses. One subject (from the knowledgeable group) with fewer

than 25 artifact-free trials per stimulus (after removal of both EOG and EEG artifacts) was removed from the final analysis. Thus, data were analyzed from only 12 subjects in the knowledgeable group.

Analysis methods

P300 amplitude was measured using the peak-peak method (Meijer, Smulders, Merckelbach, & Wolf, 2007; Soskins, Rosenfeld, & Niendam, 2001). Our algorithm searched (a) a window from 300 ms to 650 ms to find the maximally positive segment of 100 ms and (b) a window from the midpoint of that maximally positive segment out to 1,300 ms to find the maximally negative segment of 100 ms. The peak-peak amplitude of the P300 was defined as the difference in amplitude between these two segments. We used this peak-peak method rather than the more traditional base-peak method (which measures the P300 amplitude by comparing the positive peak with the prestimulus baseline amplitude) because both Soskins et al. (2001) and Meijer et al. (2007) have found the peak-peak method at least 25% more accurate in detection of concealed information. ERP analysis was performed only on the half of the trial that featured probe and irrelevant items, and not on the target/nontarget task.

Within-subjects bootstrap analysis

To determine whether the P300 component evoked by a given stimulus was greater than that evoked by another stimulus within an individual in each block, we used the bootstrap method (Wasserman & Bockenholt, 1989) at the Pz site, where the P300 amplitude is usually largest (Fabiani, Gratton, Karis, & Donchin, 1987). Because the actual distributions of average amplitudes in response to probe and irrelevant items were not available, they had to be generated by bootstrapping from the existing data. To do this, a computer program drew, with replacement, a set of single-trial probe waveforms that was equal in number to the number of accepted trials containing probe items in each block and also drew, with replacement, an equal number of single-trial waveforms in response to irrelevant items, selected randomly from among all five irrelevant items in each block. Thus, if an individual subject's block contained 35 accepted trials with probe items, the program would draw at random 35 of those trials (with replacement) and then draw at random 35 trials containing irrelevant items (with replacement).

The program then determined the mean amplitude of each set of trials and subtracted the mean P300 amplitude in response to irrelevant items from the mean P300 amplitude in response to probe items. This process was repeated 1,000 times to create a distribution of bootstrapped averages. In reporting bootstrap values, we report the number of iterations (out of 1,000) in which the average for the probe item exceeded the average for the irrelevant items in each of the three blocks. Thus, if a subject's bootstrap score on a given block were 950, that means in 950 out of the 1,000 bootstrap iterations, the P300 amplitude for probe items was greater than the P300 amplitude for irrelevant items. Our measure of detection was the average number of iterations in which the probe-items average exceeded the irrelevant-items average across all three blocks. This bootstrapping procedure was conducted for each block, and each subject's three blocks were then averaged to yield the subject's bootstrap value across blocks, as seen in Table 2. The maximum bootstrap value per block was 1,000, and thus the maximum average value per subject was 3,000/3 = 1,000. We conducted two separate types of bootstrap test: one comparing the probe item with the average of all irrelevant items in the block (Iall) and one comparing the probe item with the irrelevant item that had the largest P300 amplitude (Imax).

Results

All p values reported for within-subjects analyses of variance (ANOVAs) are Greenhouse-Geisser corrected if sphericity was violated, as indicated by a significant value on Mauchly's test of sphericity. Partial-eta squared values or Cohen's d values are reported where applicable.

Group ERP data

Figure 2 shows grand-average waveforms for all three blocks across all subjects at site Pz. A 3 (block: 1 vs. 2 vs. 3) × 2 (stimulus: probe vs. irrelevant) × 2 (group: knowledgeable vs. nonknowledgeable) mixed-model ANOVA was conducted on the peak-peak P300 amplitudes (see Figs. 3 and 4). There was a significant main effect of stimulus, with P300 amplitudes for probe items (M = 6.95 μV, 95% confidence interval, or CI = [5.16, 8.74]) larger than P300 amplitudes for irrelevant items (M = 4.12 μV, 95% CI = [3.16, 5.08]), $F(1, 22) = 47.96$, $p < .001$, $\eta_p^2 = .686$. There was a nonsignificant tendency toward a main effect of group, with knowledgeable subjects (M = 6.84 μV, 95% CI = [4.80, 8.78]) generating moderately larger P300 amplitudes than nonknowledgeable subjects (M = 4.23 μV, 95% CI = [3.59, 4.87]), $F(1, 22) = 3.07$, $p = .094$, $\eta_p^2 = .123$. There was a trend toward a main effect of block, $F(2, 44) = 2.73$, $p = .076$, $\eta_p^2 = .111$. The stimulus-by-group interaction was significant, $F(1, 22) = 35.43$, $p < .001$, $\eta_p^2 = .617$. There was a trend toward an interaction between stimulus and block, $F(2, 44) = 2.918$, $p = .066$, $\eta_p^2 = .117$. The three-way interaction was also significant, $F(2, 44) = 4.46$, $p = .017$, $\eta_p^2 = .168$.

Table 2. Individual Subjects' Bootstrap Data

	Iall		Imax	
Subject	Knowledgeable group	Nonknowledgeable group	Knowledgeable group	Nonknowledgeable group
1	991	703	721	195
2	933	653	509	343
3	967	686	876	292
4	965	631	989	401
5	876	753	537	467
6	998	103	960	77
7	865	676	641	309
8	849	478	725	118
9	1,000	586	996	311
10	907	472	729	274
11	934	493	923	271
12	962	706	849	407
Mean	937	578	788	289

Note: The table shows the average number of iterations (across all three blocks) in which P300 amplitude in response to the probe item was greater than P300 amplitude in response to all irrelevant items (Iall) and to the irrelevant item that elicited the largest P300 amplitude (Imax). The area under the receiver-operating-characteristic curve was 1.0 for the two analyses.

To decompose the stimulus-by-group interaction, we conducted two paired-samples *t* tests comparing P300 amplitudes in response to probe and irrelevant items for the knowledgeable group and the nonknowledgeable group separately. For the knowledgeable group, amplitude for probe items (M = 9.48, 95% CI = [7.46, 11.50]) was larger than amplitude for irrelevant items (M = 4.02 μV, 95% CI = [2.80, 5.24]), $t(11) = 6.87$, $p < .001$, $d = 1.07$. For the nonknowledgeable group, amplitude for probe items (M = 4.43 μV, 95% CI = [3.85, 5.01]) did not differ significantly from amplitude for irrelevant items (M = 4.04, 95% CI = [3.58, 4.50]), $t(11) = 1.41$, $p = .187$, $d = 0.24$, which indicates that the stimulus-by-group interaction was driven by the difference in amplitude between probe and irrelevant items in the knowledgeable group (see Fig. 5).

To decompose the three-way interaction, we conducted two separate 3 (block: 1 vs. 2 vs. 3) × 2 (stimulus: probe vs. irrelevant) ANOVAs, one for each group. The stimulus-by-block interaction was significant only

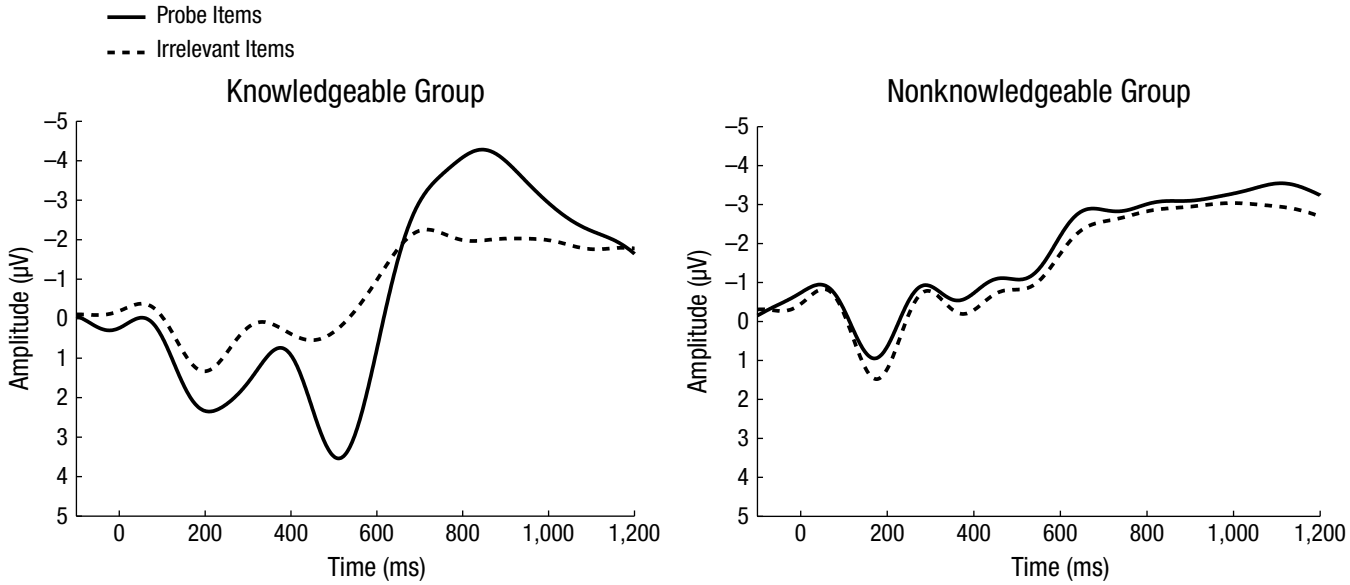

Fig. 2. Grand-average event-related-potential waveforms at site Pz for probe and irrelevant items across all three blocks, separately for the knowledgeable and nonknowledgeable groups.

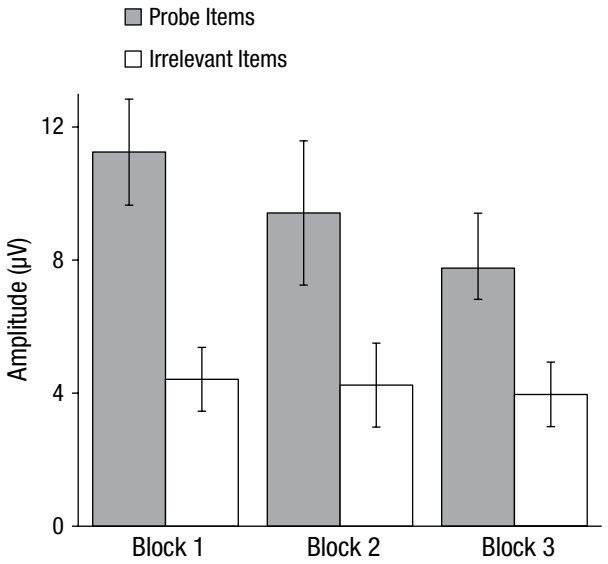

Fig. 3. Mean P300 amplitude in the knowledgeable group as a function of block and item type. Error bars show standard errors.

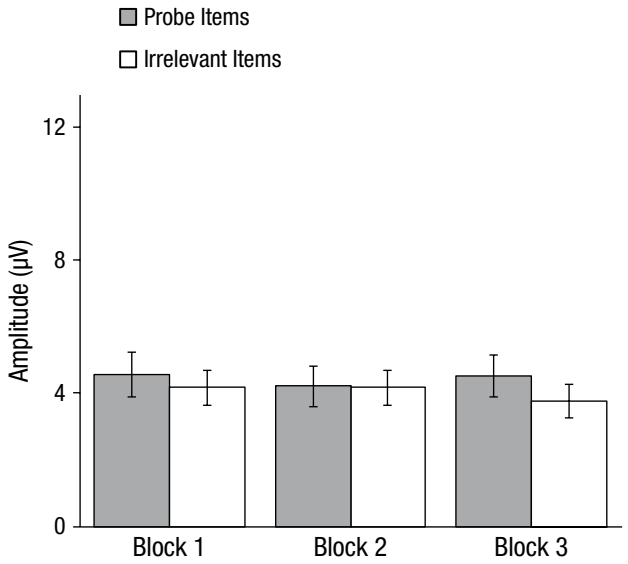

Fig. 4. Mean P300 amplitude in the nonknowledgeable group as a function of block and item type. Error bars show standard errors.

in the knowledgeable group, $F(2, 22) = 4.06$, $p = .032$, $\eta_p^2 = .270$. To decompose the stimulus-by-block interaction in the knowledgeable group, we conducted two repeated measures, one-way ANOVAs with block as the factor, one for probe items and one for irrelevant items. There was a significant effect of block only on the P300 amplitude for probe items in the knowledgeable group, $F(2, 22) = 4.556$, $p = .022$, $\eta_p^2 = .293$, which indicates that the stimulus-by-block interaction was driven by P300 amplitude differences in response to probe items across blocks in the knowledgeable group. Tukey post hoc tests revealed that P300 amplitude for probe items in this group was larger in Block 1 ($M = 11.25$, 95% CI = [9.0, 13.5]) than in Block 3 ($M = 7.76$, 95% CI = [6.78, 8.74]), $p = .007$.[2]

Individual detection efficiency

Table 2 shows the average number of iterations in the bootstrap test (out of the maximum possible 1,000) in which amplitude in response to the probe item exceeded (a) amplitudes in response to all irrelevant items (Iall) and (b) amplitude in response to the irrelevant item that elicited the largest P300 amplitude (Imax). Results are shown collapsed across blocks for each subject in the experiment, using both the Iall and Imax tests.

To examine the overall criterion-independent group discrimination efficiency of each analysis method, we conducted receiver-operating-characteristic (ROC) analyses (Green & Swets, 1966). The input statistic for the ROC analysis was the bootstrap score displayed in Table 2. Because there was no overlap between the knowledgeable and nonknowledgeable groups (see Table 2 and Fig. 6) for either the Iall or Imax analysis method, the area under the curve (AUC) was 1.0 for both methods. ROC analyses were also conducted using the bootstrap value obtained from each individual block (rather than averaged for each subject; see Table 2 and Fig. 7) as a measure of the group discrimination efficiency of single

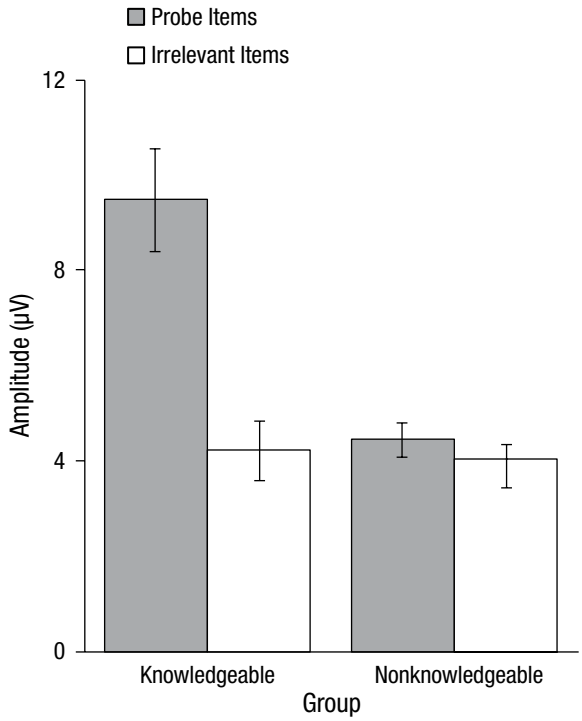

Fig. 5. Mean P300 amplitude for probe and irrelevant items as a function of group. Error bars show standard errors.

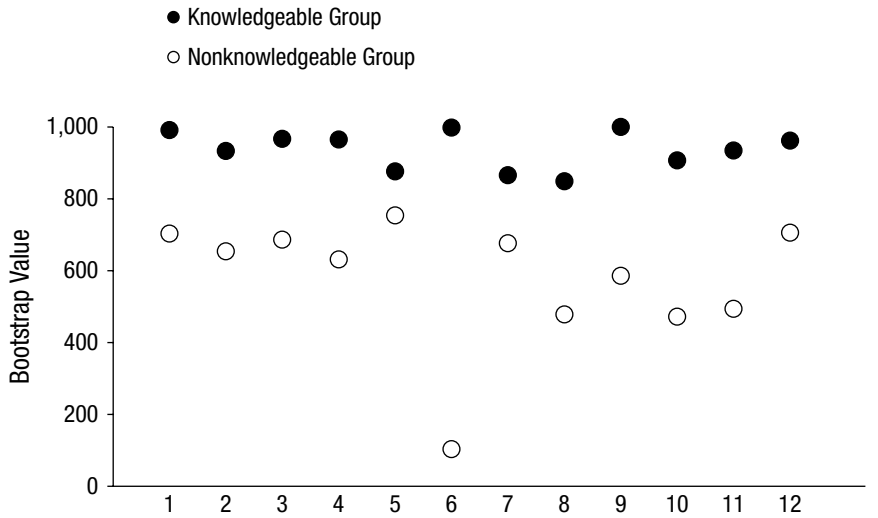

Fig. 6. Bootstrap results by individual. The graph shows, for each subject in the two groups, the number of bootstrap iterations in which the average P300 amplitude in response to probe items exceeded the average P300 amplitude in response to all irrelevant items (Iall). The area under the curve was equal to 1.0.

items. We did not compute detection rates for these analyses, as in order to do so, we would have needed to develop reasonable a priori decision criteria that are not yet available for a finalized version of the present application. Instead, to evaluate test efficiency, we relied solely on the criterion-independent, commonly accepted AUC values (but for another view of this matter, see Rosenfeld, Hu, Labkovsky, Meixner, Winograd, 2013).

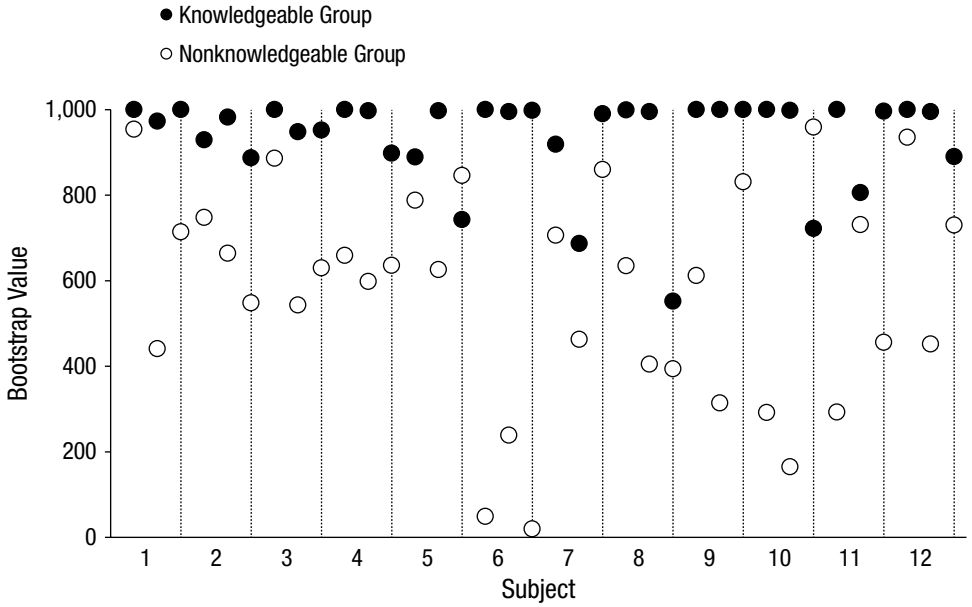

Fig. 7. Bootstrap results by subject and block. The graph shows, for each block, the number of bootstrap iterations in which the average P300 amplitude in response to probe items exceeded the average P300 amplitude in response to all irrelevant items (Iall). For each subject in the two groups, the three dots represent results for Blocks 1 through 3, with results for the third block appearing on the vertical line. The area under the curve was equal to .937.

Discussion

The data reported here demonstrate that the P300-based CIT can be highly effective in detecting real-world information acquired outside of the laboratory in a typical daily life setting. To our knowledge, this is the first study to make such a demonstration. These data may be especially relevant to researchers interested in field application of the P300-based CIT, because compared with a mock crime paradigm, this paradigm more closely approximates the conditions that would be present in a real crime scenario. Our subjects voluntarily performed a number of typical everyday acts in a real-world environment containing much distracting information, and we were able to detect information incidentally acquired through their behavior. Notably, we demonstrated that such information can be detected at the individual-subject level, which is critical for the applied use of the CIT.

In selecting probe items to be used in the CIT, we attempted to choose items most likely to be recognized and thus detected, as one would in a field CIT. Of course, it was impossible to rigorously control probe items across subjects in this design, because each subject encountered unique stimuli while wearing the camera. As Table 1 shows, we used similar categories of probe items when possible. However, as would be the case in the field, certain probe items were likely more salient to subjects than others. Further, the items we selected likely varied in the amount that they were rehearsed by knowledgeable subjects. For example, Subject 1 saw stimuli in one block that related to the color of an item she carried (likely a relatively unrehearsed detail) but saw stimuli in another block relating to the name of a professor whose course she had attended that day (likely a well-rehearsed detail). As we note in the appendix, however, we found no difference in P300 amplitudes between well-rehearsed and not-well-rehearsed probe and irrelevant items, and we thus conclude that our high rate of group discrimination efficiency is not driven by the prior familiarity of some probe items.[3]

Further, we consider the methods used here similar to those that would be used in a real crime scenario. In the field, some CIT stimuli will likely be well-rehearsed, such as the name of an accomplice or a personal object left at the scene of a crime (e.g., a cell phone), whereas other details will likely not be well-rehearsed, such as unique features of the crime scene if the perpetrator had not been there before. Additionally, the variability found here between well-rehearsed and not-well-rehearsed probe items does not threaten our present comparisons between knowledgeable and nonknowledgeable individuals. Because each nonknowledgeable subject saw stimuli identical to stimuli seen by one of the knowledgeable subjects, each nonknowledgeable subject served as a yoked control.

There are, however, several drawbacks to the present design that limit its external validity. First, although we derived memorable probe items from the 4-hr recording, such salient probe items may not always be identifiable when the only information available to investigators is a crime scene, as would often be the case in the field. Additionally, perpetrators of crimes may frequently be under the influence of alcohol or other drugs, which will likely influence the extent to which they encode memories of crime-related information.

Second, the period during which subjects created their recordings was likely made more artificial simply because of the presence of the camera itself. In viewing the videos, experimenters observed people around the subject inquiring about the camera, potentially enhancing awareness. It is unclear whether this awareness artificially inflated or deflated the accuracy of the subsequent P300-based CIT (or had no effect at all), but such a report indicates, at a minimum, that subjects were regularly aware of their participation in an experiment, which makes the task more artificial. Future studies may benefit from decreased size and obtrusiveness of the camera as technology improves.

Third, all of our participants were tested only 1 day after being exposed to the critical items in the CIT. This timescale may not be realistic in the field, where a criminal suspect may not be apprehended until long after the crime (or another event in which critical information is acquired). Several studies have addressed the influence of time delays on CIT sensitivity (e.g, Carmel et al., 2003; Gamer, Kosiol, & Vossel, 2010; Hu & Rosenfeld, 2012; Nahari & Ben-Shakhar, 2011; Peth, Vossel, & Gamer, 2012). Such a follow-up using the present protocol would be valuable as well.

Fourth, as noted previously, subjects' level of emotional arousal during encoding of probe information in the current experiment was likely quite different than the level of emotional arousal that would be experienced by a perpetrator during a crime. It is well-established that when attention is limited, emotionally meaningful stimuli tend to capture attention more easily than less meaningful stimuli (Fox, Russo, Bowles, & Dutton, 2001; Öhman, Flykt, & Esteves, 2001). This implies that emotionally meaningful stimuli encountered during a real crime (e.g., the face of a victim) might capture more attention than the stimuli encoded during the video recording in the present research (which were relatively unemotional in nature), and this could potentially lead to greater group discrimination efficiency. Additionally, events that trigger emotional responses are more likely to be remembered over time as compared with nonemotional events (Phelps, 2004), which may mean that real-world crime stimuli may

continue to capture attention following a delay compared with more arbitrarily selected stimuli, such as the ones used here. Indeed, one recent study demonstrated that emotional arousal during the commission of a mock crime reduced the extent to which a delay affects a polygraph-based CIT, though emotional arousal in that study also reduced memory for information that was not central to the crime (Peth et al., 2012). Because the lack of emotional arousal in the typical lab context suggests that lab studies might underestimate the sensitivity of the P300-based CIT, continued systematic study of how emotional stimuli affect the P300-based CIT would be highly useful.

Appendix

We attempted to, in part, ensure that there were not large differences in detection sensitivity based on the level of likely rehearsal or familiarity of probe items to knowledgeable subjects. To do this, we identified all stimuli that were likely rehearsed prior to the Day 1 camera-wearing session and determined that 12 items were likely to have been previously familiar to subjects. They can be split into roughly three categories (see Table 1 in the main text): (a) names of individuals that the subject knew prior to the camera-wearing session (Subject 1, Block 3; Subject 3, Block 2; Subject 7, Block 1; Subject 10, Block 2; Subject 11, Block 1), (b) names of classes that the subject was previously exposed to (Subject 1, Block 1; Subject 2, Block 3; Subject 6, Block 1), and (c) names of companies or schools that the subject had likely been regularly exposed to in the past (Subject 4, Block 1; Subject 5, Block 3; Subject 6, Block 3).

We conducted a *t* test comparing the P300 difference between probe items and irrelevant items for familiar items ($N = 12$) and unfamiliar items ($N = 24$). For unfamiliar items, the numerical difference between probe and irrelevant items was slightly larger ($M = 5.66$ μV) than it was for familiar items ($M = 5.41$ μV), though the difference was far from significant, $t(34) = 0.2$, $p = .84$. Further, our group discrimination efficiency did not change even if we removed blocks containing familiar stimuli from the analysis entirely. Ten of our 12 subjects had at least two blocks of unfamiliar stimuli. Thus, we conducted bootstrap tests using the two unfamiliar stimulus blocks for Subjects 2 through 5, 7, and 10 through 12, and two randomly selected blocks for Subjects 8 (Blocks 2 and 3) and 9 (Blocks 1 and 3). We also likewise conducted bootstrap tests for the corresponding blocks from the nonknowledgeable group. Table A1 and Figure A1 show the number of iterations in the bootstrap test for which the average amplitude in response to probe items exceeded the average amplitude in response to all irrelevant items (Iall) out of the maximum possible 1,000 for this test. As when familiar stimuli were included, we found no overlap between the knowledgeable and nonknowledgeable groups—the AUC for the ROC remained 1.0.

Table A1. Bootstrap Data by Group for Blocks With Unfamiliar Probe Items: Average Number of Iterations in Which P300 Amplitude in Response to the Probe Item Was Greater Than P300 Amplitude in Response to All Irrelevant Items

	Group	
Subject	Knowledgeable	Nonknowledgeable
2	955.5	706
3	976	758
4	947.5	617
5	943	707
7	838.5	661.5
8	997	520
9	1,000	463
10	861	625.5
11	901	593.5
12	942.5	591
Mean	936.2	624.25

Note: The area under the curve was equal to 1.0.

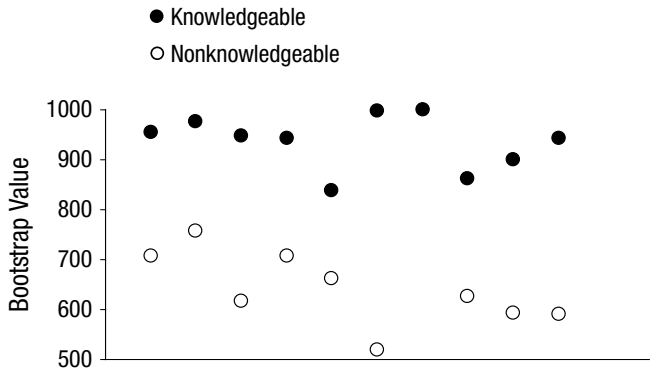

Fig. A1. Bootstrap results by individual for blocks with unfamiliar probe items. The graph shows results for the analysis in which the average amplitude in response to the probe item exceeded the average amplitude in response to all irrelevant items (Iall).

Author Contributions

J. B. Meixner designed and conducted the study and drafted the manuscript. J. P. Rosenfeld aided in design of the study and editing of the manuscript.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Notes

1. We used only three electrode sites because our aim was to determine how successfully one can use the P300 component

(our primary dependent variable) to detect incidentally acquired information. Because this component is largest at Pz, there was no benefit to using a denser array of electrodes. We recorded from sites Cz and Fz primarily to ensure that each subject demonstrated the parietal-central scalp distribution associated with the P300.

2. We considered the possibility that this order effect could be driven by fatigue, which might be manifested by an increase in blinking and other artifact-producing movements across blocks. To test that theory, we conducted a 2 (group) × 3 (block) repeated measures ANOVA, using the average number of trials per stimulus, per block, that were not lost because of artifacts. There was no main effect of block, $F(1, 22) = 0.508$, $p = .484$, and no main effect of group, $F(1, 22) = 0.004$, $p = .95$. Further, the interaction was not significant ($p = .233$).

3. In the past, we have found that highly recognizable self-referring stimuli, such as the subject's birth date, yield greater P300 amplitude than incidentally acquired stimuli when used as probe items in a P300-based CIT (e.g., Rosenfeld, Biroschak, & Furedy, 2006). We did not see a similar difference in the present study between probe items that we expected were previously rehearsed and those that were not. We suspect that the more well-rehearsed probe items in the present study were simply not as salient as the self-referring information used in our previous work, which is rehearsed repeatedly over the course of a lifetime. Thus, it is unsurprising that our items did not differ in terms of P300 amplitude.

References

Ben-Shakhar, G., Bar-Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the guilty knowledge technique in court. *Law and Human Behavior*, *26*, 527–541.

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

Cabeza, R., Prince, S. E., Daselaar, S. M., Greenburg, D. L., Budde, M., Dolcos, F., . . . Rubin, D. C. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: An fMRI study using a novel photo paradigm. *Journal of Cognitive Neuroscience*, *16*, 1583–1594.

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.

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.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*, 681–700.

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biological Psychology*, *83*, 101–107.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: John Wiley & Sons.

Hodges, S., Berry, E., & Wood, K. (2011). SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*, *19*, 685–696.

Hu, X., & Rosenfeld, J. P. (2012). Combining the P300-complex trial-based Concealed Information Test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection. *Psychophysiology*, *49*, 1090–1100.

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

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.

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.

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology*, *48*, 149–154.

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.

Milton, F., Muhlert, N., Butler, C. R., Smith, A., Benattayallah, A., & Zeman, A. M. (2011). An fMRI study of long-term everyday memory using SenseCam. *Memory*, *19*, 733–744.

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.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466–478.

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.

Phelps, E. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*, 198–202.

Rosenfeld, J. P. (2006). *The complex trial protocol: A novel, P300-based method for detection of deception*. Unpublished manuscript, Department of Psychology, Northwestern University, Evanston, IL.

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology*, *60*, 251–259.

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*, *42*, 157–161.

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*, 118–134.

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.

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: Complex vs. simple paradigms. *International Journal of Psychophysiology*, *40*, 173–180.

St. Jacques, P. L., Conway, M. A., Lowder, M. W., & Cabeza, R. (2010). Watching my mind unfold versus yours: An fMRI study using a novel camera technology to examine neural differences in self-projection of self versus other perspectives. *Journal of Cognitive Neuroscience*, *23*, 1275–1284.

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

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

Winograd, M. R., & Rosenfeld, J. P. (2011). Mock crime application of the Complex Trial Protocol (CTP) P300-based concealed information test. *Psychophysiology*, *48*, 155–161.