Received: 10 December 2018 | Revised: 21 February 2019 | Accepted: 21 February 2019

DOI: 10.1111/psyp.13362

SPECIAL ISSUE: FIFTY YEARS OF P300: WHERE ARE WE NOW?

WILEY | PSYCHOPHYSIOLOGY

P300 in detecting concealed information and deception: A review

J. Peter Rosenfeld

Department of Psychology, Northwestern University, Evanston, Illinois

Correspondence
J. Peter Rosenfeld, Cresap Lab 206, Department of Psychology, Northwestern University, Evanston, IL 60208.
Email: jp-rosenfeld@northwestern.edu

Abstract

Studies using the P300 ERP in detection of concealed information are reviewed. An overview of the initial findings, methodological issues, and use of bootstrapping methods for data analysis are considered, with various protocols explicated. Applications to forensic issues, employee screening, cognitive deficit malingering, and facial recognition in lineups are outlined. Countermeasures to the original P300-based tests are described, and a possible approach to this problem using a new complex-trial protocol is offered. Applications of this protocol to forensic and antiterror scenarios are then presented, along with its first independent replication. Studies of visual versus auditory stimulus presentation in the complex-trial protocol are evaluated. Findings from attempted voluntary suppression of P300 as a recognition signal are presented, and the effects of motivational manipulations on the P300-based complex-trial protocol are summarized. Limitations of the research are reviewed, and, based on this review, future directions of P300 methods for detection of concealed information and deception are suggested that may guide the development of more precision and reliability of this promising tool.

KEYWORDS

Concealed Information Test, deception, P300

1 | INTRODUCTION

The P300 ERP is a positive deflection of the EEG that occurs approximately 300–800 ms following presentation of a rare and meaningful stimulus within a series also including frequently presented, less meaningful stimuli (Johnson, 1986; Polich, 2007). Karis, Fabiani, and Donchin (1984) reported that participants initially exposed to word lists and later asked to discriminate between 60 of these older words and 60 new words showed larger P300s to the correctly recalled previously seen words than to correctly recalled new words. The P300 elicited by the old words may have reflected better remembered information, suggesting that this protocol could be modified and adapted to assay crime-related knowledge for use in a Guilty Knowledge or Concealed Information Test (CIT; Rosenfeld et al., 1988).

2 | EARLY STUDIES OF THE P300 CIT

The original CIT was invented by Lykken (1959). Two kinds of items were presented: critical probe items, such as a murder weapon related to an incident under investigation, and noncritical irrelevant items unrelated to the incident but in the same category as the critical items, such as other weapons. This CIT method measured the skin conductance response as a sign of autonomic nervous system (ANS) arousal probably related to the orienting reflex (klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2015, 2017). Although the ANS-based CIT was reasonably accurate (90% sensitivity, 100% specificity), the ANS may have been only indirectly responsive to specific stimuli, whereas the P300 might have provided a more accurate index of recognition. This led to the first full-length published studies of the P300-based

CIT (Rosenfeld et al., 1988; Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987). Here, each student participant pretended to steal a small item from a black-painted lunch box. There were 10 items inside—a $10 bill, a wrist watch, a portable radio, and so on. The selected probe item was placed in a pocket. Then, the participant was told he would be given a brain wave lie detector test: He was to watch a display screen on which various item names would appear about once every 3 s. He was told to press a *no* button to all items except one predesignated irrelevant item that had not been in the box. This was a test target item that forced his attention to all items since he did not know which item would be presented on a given CIT test trial. He was also told that the item he pretended to steal would be presented on occasion but that he was to lie and press *no* indicating nonrecognition whenever it appeared. He was also told that, even though he pressed *no*, it was expected that his brain wave response upon seeing the stolen item would give him away. The results were that, in 9 of 10 participants, a P300 component was elicited in response to the probe item but not to the irrelevant items. The reported 90% sensitivity (there were no unknowledgeable subjects run to allow determination of specificity) was about the same as that reported by Lykken (1959). In subsequent studies, the autonomic CIT typically reported 80%–90% accuracy with sensitivity and sensibility combined, whereas the P300 CIT yielded slightly higher overall accuracies, 85%–100%. Additional support was found in a meta-analysis with effect sizes (Cohen's *d*) of 1.89 for the P300 and 1.55 for skin conductance (Meijer, Selle, Elber, & Ben-Shakhar, 2014). These results showed that P300 signaled involuntary recognition of a denied mock crime detail, the essence of a CIT. (Many other recent methods of concealed information and deception detection are reviewed in the volume edited by Rosenfeld, 2018).

Farwell and Donchin (1991) employed P300 in a CIT testing four previously admitted wrongdoers at the University of Illinois. The probe stimuli were items relevant to a minor crime under investigation, such as the place of the crime scene or the name of another person involved. For each of the six probe stimuli, there were one target and four irrelevant stimuli. The target and irrelevant stimuli were items of the same type as the probe stimuli, such as a location different from the actual crime scene or a fictitious name. Participants counted the predesignated irrelevant targets but were instructed to ignore probes and irrelevants. Probe items but not irrelevants elicited P300. It was concluded that P300 was a good guilty knowledge index.

However, a problem with this study was the question of whether the P300 indexed recognition of crime details versus the recognition of well-rehearsed facts admitted after the crime but prior to the CIT due to interrogations by parents, administrators, and campus security officials (Rosenfeld, 2005, 2011). The same issue can be raised about the main study also reported in which student participants committed mock espionage acts whose details were rehearsed with the experimenters to a standard of perfection before the P300 test (Farwell & Donchin, 1991).

In another study, we assessed student participants with P300 methods about antisocial/illegal acts from their past, such as cheating on tests, plagiarizing papers, using false IDs, and so on (Rosenfeld, Angell, Johnson, & Qian, 1991). Experimenters did ask participants prior to the testing to privately examine a list of antisocial acts and to check boxes next to the list items that applied to them. They did this only once and retained their lists; experimenters secretly recorded their answers with a hidden video camera. However, it remained unclear whether the subsequent P300 signs of their guilty acts would have occurred had experimenters not used the checklists. Filling out the checklists constituted a rehearsal, such that the findings were not ecologically valid.

This difficulty was next addressed by testing participants on only one item designated without using any pretest checklist (Johnson & Rosenfeld, 1992). This item was about cheating on tests, since about half the participants in the previous study privately acknowledged cheating on tests (Rosenfeld et al., 1991). It was correctly assumed that assessing a new sample from this same population based on this same cheating item would provide guilty and innocent groups of approximately equal sizes. The real truth was verified by giving the checklists after the P300 CIT in conditions of perceived privacy. The resulting overall diagnostic accuracy combining sensitivity and specificity was about 87% (Johnson & Rosenfeld, 1992). In the next section, the methods for determining these diagnostic accuracies are detailed.

3 | DIAGNOSTIC AND DEFINITIONAL ISSUES

The method used to determine diagnostic accuracy in the early P300 CIT studies was the bootstrap (Efron & Tibshirani, 1994; Farwell & Donchin, 1991). P300 CITs employed three kinds of stimuli: the probe (P), irrelevant (I), and target (T). This protocol was therefore named the three-stimulus protocol (3ST; Rosenfeld, Biroschak, & Furedy, 2006). Probes were items that were directly relevant to the information sought, such as the specific murder weapon (e.g., .357 magnum) in a crime. Other items from the same category were presented as irrelevants (e.g., .38 colt, .45 automatic, .22 Beretta, etc.). Probes were infrequently presented (e.g., $p = 0.15$), whereas irrelevants were frequent ($p = 0.70$). Probe and irrelevant items correspond exactly to critical and noncritical items in Lykken (1959). The remaining stimuli presented were targets ($p = 0.15$).These were irrelevant stimuli but required a unique response. Participants were instructed to press a right mouse button to recognized targets but the left mouse button to other items: probes and nontarget irrelevants. Probes and targets were rare and meaningful for guilty participants, rareness and meaningfulness being the antecedent conditions for P300 elicitation (Johnson, 1986). Both these stimuli were therefore expected to elicit P300s in guilty or knowledgeable participants. The frequent nontarget irrelevants were expected to elicit small or absent P300s.

There have been two variations of the 3ST: (a) the single probe protocol in which, in each block of trials, there is one probe, one target, and four to seven different irrelevants, each item repeated 20–40 times (Rosenfeld et al., 1991, 1988); and (b) the multiple probe protocol consisting of several differing probe, irrelevant, and target stimuli, in the similar ratio of 1:4:1, each item repeated less often (Farwell & Donchin, 1991). It was found that the multiple probe protocol is more demanding than the single probe version, as indexed by increased reaction time (RT) to all stimuli, and that, consequently, the single probe protocol is more accurate (Rosenfeld, Shue, & Singer, 2007).

The approach to individual diagnosis measured the "CIT effect" (Rosenfeld, 2011). This is the difference between the physiological indices in response to the average probe versus irrelevant stimulus (klein Selle et al., 2015, 2017). In this situation, the target stimulus is used only as an attention holder: The participant must attend to all the stimuli in order to not miss the targets, responses to which constitute the explicit task. If s/he misses targets, the participant is not processing the stimuli, and the results are discarded. The bootstrap technique involves repeated resampling with replacement multiple times from a participant's set of single sweeps of probe and irrelevant P300 amplitudes (Rosenfeld & Donchin, 2015; Rosenfeld, Ward, Meijer, & Yukhnenko, 2017). Probe-minus-irrelevant amplitude differences are computed for each resampling, and the process is repeated 100 times. If 90% or more of these iterated mean differences are positive (probe > irrelevant), then the participant is diagnosed as knowledgeable of the probe, from which guilt may be inferred. Typically, this rule implies another: if the participant does not show 90% or more probe > irrelevant P300 differences, the participant is classified as innocent (Rosenfeld, 2011; Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013). Other classification algorithms are possible as described below. For the innocent or unknowledgeable participant, the probe is not known; it is simply another irrelevant stimulus, so the expected proportion of probe > irrelevant differences over many iterations is 50%. This method was called the bootstrapped amplitude difference method (BAD; Soskins, Rosenfeld, & Niendam, 2001).

A different bootstrap computation was also used (Farwell & Donchin, 1991). This procedure assumed that the task relevant and rare target stimuli should elicit P300 in all participants, as should the probe only in knowledgeable participants. Therefore the probe-target cross-correlation across most of the recording epoch will be high in knowledgeable participants. In contrast, in the unknowledgeable participant, the probe is just another irrelevant so the probe-irrelevant correlation will be larger. The sets of probe, target, and irrelevant single sweeps are each resampled with replacement multiple times, then averaged, and the probe-irrelevant cross-correlations and probe-target cross-correlations over trials are computed. This process is iterated 100 times. If 90% of comparisons show the probe-target correlation > probe-irrelevant correlation, the participant is determined to be knowledgeable. Alternatively, if 90% of the iterated comparisons show probe-irrelevant > probe-target correlation, then the participant is declared unknowledgeable. Participants whose comparisons yield neither knowledgeable nor unknowledgeable decisions are indeterminate. This method was called the bootstrapped correlation analysis of disparity (BC-AD; Soskins et al., 2001). Its weaknesses relate to occurrences of probe and target P300s with significantly different P300 peak latencies that therefore cannot show strong cross-correlations (Rosenfeld, 2011).

Other early P300 CIT studies focused on detecting concealed knowledge of newly learned verbal materials (Allen, Iacono, & Danielson, 1992). In these studies, a highly original Bayesian analysis showed outstanding accuracy in individual diagnosis. Other studies utilized a wavelet classifier method also addressed to the diagnostic challenge (Abootalebi, Moradi, & Khalilzadeh, 2006). There has never been a convincing comparison of the various methods, because the extant comparative studies never systematically compared studies all using similar threshold criteria, nor definitions of P300 measurement (e.g., base-to-peak vs. peak-to-peak, and so on; Rosenfeld, 2011). There is presently no general agreement about the best methods of P300 measurement and analysis for CIT purposes, although many investigators have settled on use of the peak-to-peak measure with bootstrap analysis of amplitude differences (Rosenfeld & Donchin, 2015; Rosenfeld et al., 2013).

We studied the prestimulus baseline-to-peak Pz P300 recording at a high-pass filter setting of .01 Hz versus the peak-to-peak recording seen at a more typical filter setting of .30 Hz in a two-stimulus oddball paradigm with rare targets and frequent nontargets (Soskins et al., 2001, extending Duncan-Johnson & Donchin, 1979). In this study, the P300 recording from the Pz electrode had its output divided into two leads: one passed through an amplifier with a .01 high-pass filter and the other passed through an amplifier with high-pass filter set to .30 Hz. At the former setting, one sees the superimposed target and nontarget P300s in the center panel of Figure 1. The .30 Hz filtered recordings are seen in the top and bottom panels of the figure at Fz and Pz,

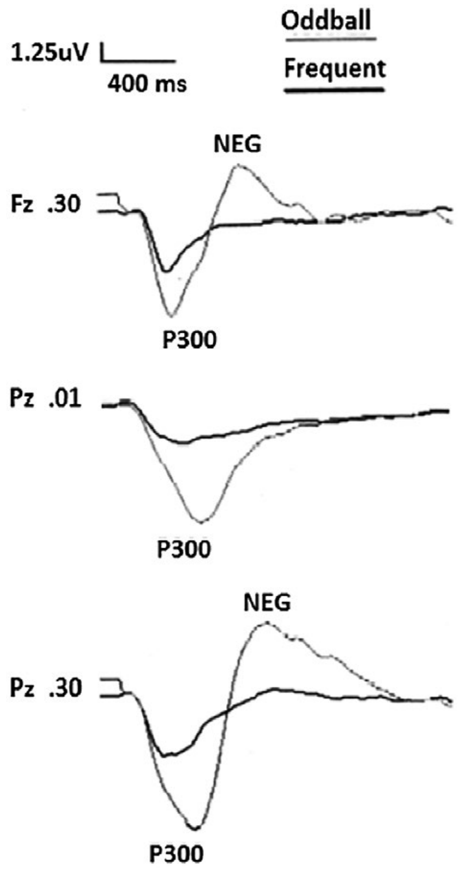

FIGURE 1 Recordings of grand-averaged P300s (from 12 participants) to frequent ($p = 0.8$) and rare/oddball ($p = 0.2$) verbal stimuli, recorded at high-pass filter settings of .01 (center) and .30 (top and bottom), the bottom two from the same Pz location and the top wave form from Fz at .30 Hz. P300s are indicated below wave forms; NEG above wave forms indicates the apparently negative component following P300 seen in top and bottom. Adapted from Soskins et al. (2001)

respectively. P300 is printed below the wave forms near the ERP peak in each case.

The greater capacitive coupling of the recording seen in the lower panel of Figure 1 (Pz at .30 Hz) causes the recovery slope of P300 to return to baseline and continue into an apparent negative component, NEG (Figure 1, above) indicated in the top and bottom wave pairs near their peaks. The amplitude of this late negative wave seen in the .30 Hz Pz channel correlated at + .67 with the Pz P300 recovery slope recorded with the .01 Hz filter setting as illustrated (Figure 1, center). P300 amplitude and its recovery slope are orthogonal, since trailing edges of the P300 recovery slopes can vary over trials whose P300 peaks have the same peak amplitude. It was hypothesized that the recovery slope of the P300 recorded at .01 Hz—or its correlate, the late negative wave in the .30 Hz channel—may thus provide information orthogonal and in addition to the peak amplitude of P300. Thus, we advocate the peak-to-peak[1] recording as the better P300 index in the P300-based CIT than base-to-peak recording, and we have reported that the former index is 25%–35% more accurate than base-to-peak recording (Rosenfeld, 2011). This finding has been replicated independently (Cutmore, Djakovic, Kebell, & Shum, 2009; Lukács et al., 2016; Meijer, Smulders, Merckelbach, & Wolf, 2007).

4 | EARLY APPLICATIONS OF P300

The early P300-based CITs mostly used forensic, employee screening, and espionage scenarios (for a review, see Rosenfeld, 2011). However, it was clear that the ability to detect denied knowledge with P300 could also be useful in the detection of malingered cognitive deficit (Ellwanger, Rosenfeld, Sweet, & Bhatt, 1996; Rosenfeld, Ellwanger, & Sweet, 1995). These workers tested university students instructed to simulate malingering by denying autobiographical knowledge such as birth dates, phone numbers, and mothers' maiden names, appearing as stimuli in an autobiographical oddball paradigm. Although, as instructed, participants behaviorally denied recognition of these self-referring items, they were all potent elicitors of P300, suggesting actual recognition. Patients with true head injury have smaller than normal P300s; suspected malingerers from this population might lack the requisite P300 ERP required by a P300-based test. It was confirmed, however, that although P300s are reduced in head injury patients, there was nevertheless a clear difference in their P300s in response to recognized versus unrecognized information (Ellwanger, Rosenfeld, & Sweet, 1997).

Others followed up this approach by applying the three-stimulus protocol to a word recognition task (Van Hooff, Brunia, & Allen, 1996; Van Hooff & Golden, 2002). Recognized previously learned nontarget words elicited P300. Since there was an interval of 1 day between learning and testing, the authors suggested that the procedure might be useful when "the integrity of memory is in question" as in malingered cognitive deficit. This approach culminated in a report in which there was a normal group performing to the best of participants' abilities and another normal sample asked to feign a credible memory disorder. ERPs as well as RT and performance on the Amsterdam Short-Term Memory test were recorded (van Hooff, Sargeant, Foster, & Schmand,

[1] Although the term peak is used in this section, the use of single peak points as amplitude values is not recommended as these can be noisy and unreliable (see Luck, 2014, pp. 284–293). All studies by the present author utilize mean amplitudes within 100-ms segments. Also, although the study by Soskins et al. (2001) recorded the NEG wave using .3 Hz high-pass filter settings, recent studies by the present author utilize a .16 Hz setting, close to the .1 Hz value urged by Tanner, Morgan-Short, and Luck (2015) as "optimal for statistical power in detecting true effects while not introducing distortions."

2009). Although memory test and RT data clearly distinguished the two groups, P300 amplitude and its scalp distribution did not. It was concluded that the discrepancy between behavioral and ERP data could be interpreted "as evidence of intentional underperformance" (malingering).

It was, however, appreciated that there were limits to the use of the autobiographical oddball paradigm in diagnosis of malingered cognitive deficit (Ellwanger, Rosenfeld, Hankin, & Sweet, 1999; Ellwanger, Tenhula, Rosenfeld, & Sweet, 2000; Rosenfeld et al., 1999, 1998; Rosenfeld, Sweet, Chuang, Ellwanger, & Song, 1996). First, most malingerers are not so unsophisticated as to verbally state that they do not recall an item such as their birth date when they may have just filled out an information card in which they provided that information. The behavioral Multidimensional Memory Test (MDMT) was developed as an entrapment procedure to catch such malingerers (Guilmette, Hart, Guiliano, & Leinenger, 1994; Hiscock & Hiscock, 1989). A sample three-digit number is presented followed 5–10 s later either by a match or mismatch, each of which has a 50% chance of appearing. Normal persons and even patients with moderate head injury perform well on this easy test, unless they are involved in litigation (Sweet, 1999). A 90% correct criterion was proposed as a cutoff for a diagnosis of malingering (Guilmette et al., 1994). In contrast, it was reported that, in a population of 15 nonlitigating closed head injury patients, two persons scored < 80% correct on the behavioral MDMT. Hence, this behavioral test and its 90% diagnostic criterion were far from infallible (Ellwanger et al., 1999, 1997).

Therefore, the MDMT was modified to include P300 data (Ellwanger et al., 1999; Ellwanger et al., 2000; Rosenfeld et al., 1996, 1998, 1999). Importantly, the relative probabilities of match and mismatch were changed from equality to 17% and 83%, respectively, and the random interval between the sample number and the subsequent test number was reduced to 3–5 s. It was found that the P300 amplitudes and scalp amplitude distributions of simulated malingerers could be readily discriminated from those of control participants, but at best the overall accuracy of individual classification never exceeded 75%. This figure applied also to later papers that introduced another modification of the MDMT in which, following the presentation of the sample number, a series of six numbers followed that included one match randomly placed in a series of mismatches (Ellwanger et al., 1999, 2000). The test was then further altered such that match and mismatch probabilities were 11% and 89%, respectively (Ellwanger et al., 1999, 2000). In this study, classification accuracies of malingering simulators and nonsimulators were .80 sensitivity and 1.0 specificity, respectively, for an overall 90% accuracy.

Another application advanced the field by applying P300 as a sign of recognition of pictorial, specifically facial identification, in the context of eyewitness testimony accuracy (LeFebvre, Marchand, Smith, & Connolly, 2007). The participants watched crimes enacted on videos in which a culprit entered the premises of a victim and stole a laptop computer. In the subsequent three-stimulus protocol test, the culprit's face was the probe, the victim's face was the attention-holding target, and five other filler faces from a photographic database and unseen by the participant—who was the "witness"—were irrelevants. The entire stimulus set modeled a suspect lineup. Participants were tested either 1 min after watching the video, 1 hr later, or 1 week later. In another condition, the actual culprit was absent from the lineup. The individual diagnostic procedure compared P300 averages converted to z scores for each stimulus versus the averages to all stimuli from each of six parietal and centroparietal scalp sites. A correct identification occurred when the z score from the actual probe exceeded that of the grand average by two or more. A misidentification occurred when one of the irrelevant faces elicited a P300 whose z score was equal to or greater than two. If none of the P300 z scores reached two, then the decision was either false rejection if there actually had been a probe present in the lineup or correct rejection if the lineup lacked a culprit.

The results were that, in the immediate and 1-hr delay conditions, 79% and 83% accuracies (sensitivity), respectively, were obtained; however, accuracy fell to 58% with a 1-week delay. The P300 for culprit present conditions was significantly greater than the largest P300 in the culprit absent condition. The authors concluded: "Taken together, these results provide convincing evidence that ERP patterns can provide a neurophysiological index of correct identification patterns." Yet the loss of accuracy after a 1-week delay suggests that this optimism should be tempered given that there are often delays between arrests and testing in the field.

5 | COUNTERMEASURES AND THE NEW COMPLEX-TRIAL PROTOCOL

Researchers appreciated that if, in the three-stimulus protocol, the experimenter designated an irrelevant stimulus as a target requiring a unique button press leading to a P300 response, then a participant could also decide to make secret responses to other irrelevants, likewise leading to P300s. The result could be no differences in the averages of probe versus irrelevant responses, thereby minimizing the diagnostic basis of the P300-based CIT (Rosenfeld, Soskins, Bosh, & Ryan, 2004). In the first experiment, participants were run in a mock crime, multiple-probe protocol in which an item of jewelry with an owner's name tag attached was stolen from a desk drawer lined with pink lining paper. The operation had an animal name ("donkey"), and there was a computer on top of the desk. Four irrelevant stimuli and one irrelevant target item were designated from each of the six probe categories:

TABLE 1 Diagnoses of guilty (knowledgeable)[a]

Method	Guilty group	Innocent group	Countermeasure group
BAD	9/11 (82%)	1/11 (9%)	2/11 (18%)
BC-AD	6/11 (54%)	0/11 (0%)	6/11 (54%)

Note. The BAD method involves bootstrapping of probe-irrelevant P300 amplitude differences so as to determine whether or not 90% of 100 difference iterations have probe > irrelevant P300 amplitude. The BC-AD method determines whether or not 90% of the probe-target P300 amplitudes are more highly correlated than probe-irrelevant P300 amplitudes. The criterion for guilty determination in both methods is 90%.
[a]In Experiment 1 of Rosenfeld et al. (2004).

jewelry type, lining paper color, owner's name, operation name, furniture type, items on top of desk. This structure was based on the multiple probe protocol used by Farwell and Donchin (1991). There were three groups: (a) simply guilty (SG), (b) guilty taught countermeasures (CM), and (c) innocent (IN), whose members stole nothing.

CM participants were taught six countermeasures, one for each of the six categories of stimulus, and executed one of these countermeasures for each irrelevant item that appeared within the same category. The countermeasures included acts such as non-noticeable right or left finger presses, toe wiggles, and mental images. Diagnostic accuracies are given in Table 1 using the BAD and BC-AD methods, both using a 90% bootstrap criterion as described above. With the BAD method, countermeasures reduce accuracies from 82% to 18%. With the BC-AD method, accuracies are poor in the first place (54%) and just as bad with countermeasures.

In the second experiment, countermeasures were taught to participants simulating malingered amnesia for their own birth dates. We used a one-probe protocol in which the stimuli were each participant's birth date, four irrelevant dates, and one designated target irrelevant date, with each date repeated 30 times. The participants were run in 3 successive weeks, in the first as simply guilty, in the second with countermeasures, and in the third with instructions to repeat the first week. The countermeasures were as before: for each of the four irrelevant dates, make either a left or right finger press or toe wiggle. The results are summarized in Table 2.

TABLE 2 Diagnoses of guilty (knowledgeable)[a]

Week/Condition	BAD method	BC-AD method
1/No countermeasures	12/13 (92%)	9/13 (69%)
2/Countermeasures	6/12 (50%)	3/12 (25%)
3/No countermeasures	7/12 (58%)	3/12 (25%)

Note. The BAD method involves bootstrapping of probe-irrelevant P300 amplitude differences so as to determine whether or not 90% of 100 difference iterations have probe > irrelevant P300 amplitude. The BC-AD method determines whether or not 90% of the probe-target P300 amplitudes are more highly correlated than probe-irrelevant P300 amplitudes. The criterion for guilty determination in both methods is 90%.
[a]In Experiment 2 of Rosenfeld et al. (2004).

Countermeasures impact this single probe protocol also, as seen with either diagnostic algorithm. The countermeasure effects persist into the third week when no countermeasures were used, as confirmed with a return of RTs in the third week from a very elevated level in the second week, when countermeasures were used, to the same levels for all stimuli as seen in the first week. In a control group in which the CIT was simply repeated for 3 weeks with no countermeasure use, the sensitivity remained above 90% in each week, suggesting that the 58% detection in the third week of the countermeasure group was not merely due to repetition. It was clear that a new countermeasure-resistant protocol was needed.

We reasoned that demand aspects of the three-stimulus protocol made it vulnerable to countermeasures since each trial contained a dual task: the explicit target-nontarget discrimination and the implicit probe-irrelevant discrimination (Rosenfeld et al., 2008). On each trial, the participant had to explicitly decide whether or not the right button target response was correct. However, it was also assumed that, when the probe or irrelevant stimulus was presented, there would also be an implicit response conflict even though both probe and irrelevant required the same left button press. Many studies have shown that a dual task situation in which one task is the P300 oddball task and the other task is orthogonal (such as flight simulation) causes a reduction in P300 amplitude from the normally higher level seen with only the oddball task (Donchin, Kramer, & Wickens, 1986). We hypothesized that P300 to the probe in the three-stimulus protocol was not as large as it could potentially be because of the division of attention between the two tasks, explicit and implicit (Rosenfeld et al., 2008). The P300 in the three-stimulus protocol is usually large enough to detect concealed knowledge in the absence of countermeasures, but, as was just shown, countermeasures pose a problem for it. So, we devised a new P300 CIT in which the two decisions were separated, as shown in Figure 2.

This figure shows that there are two stimuli presented on each trial, S1, either a probe or an irrelevant, and 1 to 2 s later, S2, either a target or nontarget. A date is shown in Figure 2 as S1 to test if a suspected malingerer can recognize the birth date. S2 shows either a target number string (11111) or a nontarget number string (22222, 33333, … 55555). The response to S1 is the same left button press on the left-hand mouse no matter which S1 is presented,

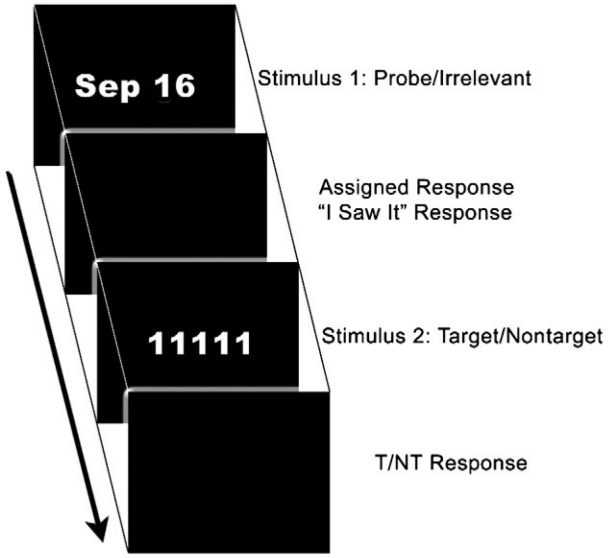

FIGURE 2 The Complex Trial Protocol event sequence, with a subject's birth date as Stimulus 1 (S1; probe or irrelevant), then the perception acknowledgement response ("I saw it"), then the target or nontarget as Stimulus 2 (S2), followed by the target/nontarget (T/NT) response. Each stimulus endures 300 ms. The S1-S2 interval randomly varies 1,300–1,800 ms

probe or irrelevant. Participants are warned in advance that there will be unannounced attention tests at random intervals on the identity of the S1, and that there will be an adverse consequence if more than one of these pop quizzes is incorrectly answered. Additionally, the S2 requires an explicit discrimination—either a left or right button press on a right-hand mouse. Two stimuli and responses on each trial led to this procedure's name: the complex-trial protocol (CTP).

In the first test of this new protocol, birthday probes were used, resulting in the data of Table 3. The design of this experiment was (a) a week with no countermeasures, (b) a week with countermeasures, and finally (c) a repeated week with no countermeasures. The diagnostic sensitivities based on the BAD test were unaffected by countermeasure use as presented in Table 3.

In this study, a unique countermeasure was performed for each of the four irrelevants. However, in subsequent studies we reported that countermeasures against any number up to five irrelevants reduced neither the sensitivity nor the specificity of the CTP below 90% (Labkovsky & Rosenfeld, 2012; Rosenfeld & Labkovsky, 2010—see Figure 3). These studies included episodic memory information as in a mock crime scenario (Winograd & Rosenfeld, 2011).

In this last cited study, participants were instructed to steal an item from a drawer in a nearby room. To maintain ecological validity, no mention of the name or nature of the stolen item was made prior to the mock crime and subsequent P300-based CTP. There were three subgroups of participants:

TABLE 3 Diagnoses of guilty (knowledgeable)[a]

Week/Condition	BAD method
1/No countermeasures	11/12 (92%)
2/Countermeasure	10/11 (91%)
3/No countermeasures	11/12 (92%)

Note. The BAD method involves bootstrapping of probe-irrelevant P300 amplitude differences so as to determine whether or not 90% of 100 difference iterations have probe > irrelevant P300 amplitude. The criterion for guilty determination is 90%.
[a]In Rosenfeld et al. (2008).

a simply guilty group, a guilty group taught countermeasures as in Rosenfeld et al. (2004), and a control innocent group whose members stole no item. Ten of 12 simply guilty subjects were accurately detected with P300. All 12 of 12 were detected in the countermeasure group. There was one false positive in 12 innocent control subjects. We also observed that only in the more cognitively challenged countermeasure group were the reaction times to probe and irrelevant stimuli elevated by 300–800 ms in comparison to a baseline condition in which all participants were innocent. So, RT again provided a useful adjunct index of attempted-countermeasure use.

A timely application of the P300-based CTP to the antiterror challenge was recently reported (Meixner & Rosenfeld, 2011). In the antiterror scenario, the aim is to prevent the criminal act by detecting it in the planning stage. One must arrest and test the suspects on not-yet-committed acts. Hence, we had simulating terrorists read informational brochures, choose the dates, places, and times of their mock acts of terror, and then write a letter to mock terrorist chieftains summarizing and justifying their attack choices. An innocent control group wrote a letter about vacation plans but took the same CTP as the guilty mock terrorists, with three separate blocks—testing on recommended date, method, and city of the planned terror act. The BAD method with 1,000 bootstrap iterations was applied to each CTP block for both groups.

The proportion of bootstrap iterations where the known probe P300 was compared with bootstrapping to the combined irrelevants P300 averaged 966/1,000 with probe greater than irrelevant in terror participants versus 546/1,000 (near chance) in controls. This yielded 100% accuracy with 12/12 correct guilt detections and 0/12 false positives in the controls, for a perfect ROC (receiver operating characteristic) area score of 1.0 (see Figure 4). But, the point of detection here is to identify suspects before the act is committed, that is, without advance knowledge of probe identity. To simulate this scenario, it was assumed that the largest P300 from a guilty participant among average P300s to all stimuli would be the average to the probe. This maximum P300 average was compared in each participant to the next largest average P300, assuming it to represent the largest response to an irrelevant

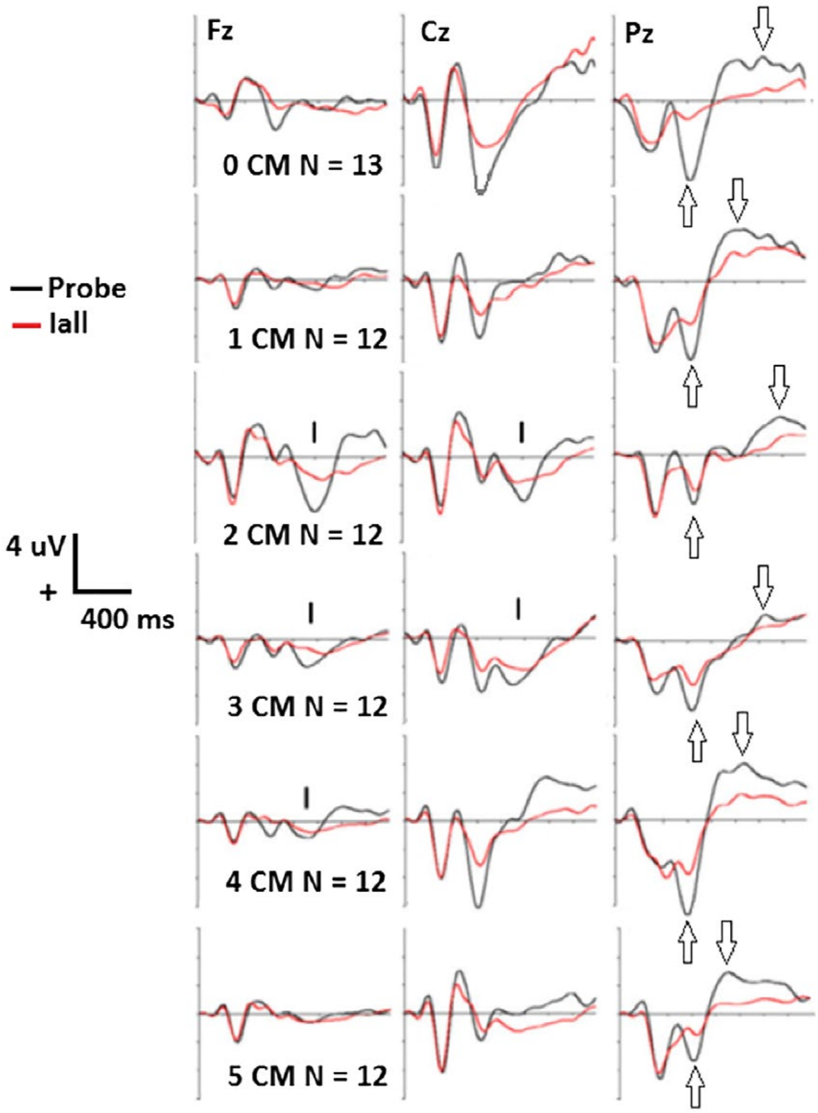

FIGURE 3 Probe and average of all irrelevant (Iall) P300s (grand averages) at Fz, Cz, and Pz in a study with 6 groups shown, one per row: A simply guilty control group with no countermeasures (0 CM), CMs against 1 irrelevant (1 CM), 2 irrelevants (2 CM) … up to all 4 irrelevants and the probe (5 CM). Arrows pointing up in Pz waves show P300s; downward arrows show negative waves on which peak-peak values are calculated. Vertical lines above *x* axes of some Fz and Cz waves indicate P900, another ERP reliably observed (from 800–1,000 ms) when CMs are performed to 40%–60% of stimuli. Adapted from Labkovsky and Rosenfeld (2012)

stimulus. That approach, called the blind Imax test, yielded 10/12 accurate detections for guilty participants with no false positives, for an area under the ROC curve = .979, a value indexing excellent discriminability. Comparing the maximum P300 to the usual average of all remaining P300 averages led to an unacceptable number of false positives.

6 | INDEPENDENT REPLICATION OF THE CTP AND ATTEMPTED SUPPRESSION

The main findings of the CTP in scenarios in which the probe is known in advance were independently replicated (Lukács et al., 2016). They found that, if the probe P300 average is compared in the bootstrap with the average of all irrelevant P300s, the sensitivity was > 90%, even in participants using the usual countermeasures in which irrelevants were made into secret targets with covert responses. However, these workers developed a novel countermeasure that significantly reduced the sensitivity of the CTP when the blind Imax test was used, as in proactive antiterrorist protocols (Meixner & Rosenfeld, 2011). The countermeasure was to have participants (a) secretly and mentally say a word like *dog* to themselves when seeing either the probe or any one of four irrelevant items, and (b) secretly and mentally say a word like *cat* when seeing the remaining two irrelevant items used. These last two irrelevants became the rare and meaningful items evoking the largest P300s, thereby defeating the blind Imax test. Our view that we are presently testing is that this countermeasure is difficult, requiring much practice.

Other countertactics have also been recently attempted against the CTP, in particular, the technique of mental suppression (Anderson & Levy, 2009). Evidence for suppression is largely based on demonstrations of suppression of response term memories in paired associate learning (Anderson & Green, 2001). Further support has come

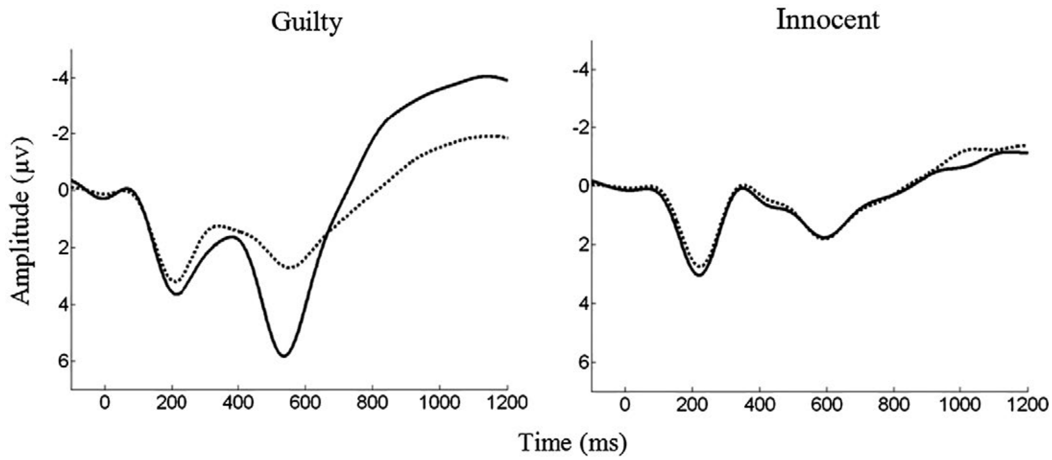

FIGURE 4 Average Pz ERPs in mock terror study: The average known probe (solid line) and irrelevant (dotted line) ERP waveforms in guilty mock terrorists (left) and innocent controls (right) adapted from Meixner and Rosenfeld (2011). Clear CIT effects (probe-irrelevant differences) are seen at left but not at right

from recent reports of ERP correlates of such suppression (Bergstrom, Anderson, Buda, Simons, & Richardson-Klavehn, 2013; Hu, Bergström, Bodenhausen, & Rosenfeld, 2015). These studies reported that the P300 sign of recognition could be voluntarily suppressed. The former report, however, used the older, CM-vulnerable three-stimulus protocol, and, as additionally shown, the claim of P300 suppression was confounded by differences in amount of rehearsal between key comparison groups (Rosenfeld, Ward, Drapekin, Labkovsky, & Tullman, 2017). The other article used a CTP but, uniquely, used a 50-50 target/no-target ratio—unlike the usual 20-80 ratio—that proved to increase task demand, probably reducing P300 amplitude (Ward & Rosenfeld, 2017). Moreover, this report showed suppression only in the base-to-peak P300. With the generally preferred peak-peak method, no suppression was seen. Most recent research unambiguously showed either no effect or a paradoxically enhancing effect of suppression instructions on P300 signs of episodic and autobiographical recognition in the CTP (Rosenfeld, Ward, Drapekin et al., 2017; Ward & Rosenfeld, 2017).

7 | MOTIVATIONAL AND MODALITY INFLUENCES ON THE CTP

A recent systematic study of the effects of motivational manipulations on the P300 in the CTP demonstrated that manipulations that produced dependent behavioral effects did not affect the P300 sign of recognition of episodic or autobiographical memories in both mock crime or simulated malingering scenarios (Rosenfeld, Labkovsky, Davydova, Ward, & Rosenfeld, 2017; Rosenfeld, Sitar, Wasserman, & Ward, 2018; Rosenfeld, Ward et al., 2018). This was important because, as recently reviewed, the typically employed ANS response of skin conductance is affected by motivational manipulations in the laboratory. It could even be the case that if motivational manipulations affect ANS responses in the low stakes laboratory situation, they may be even more effective in the higher stakes field situation, although this is not necessarily the case and needs empirical confirmation. Although motivational manipulations in the laboratory situation do not affect the P300 CIT, it remains possible that higher levels of motivational influences in field situations may affect the ERP-based test, but this too needs empirical confirmation and seems less likely given the lack of effect seen in laboratory tests.

In standard detection of deception procedures presently used in law enforcement, questions are directly posed to participants in the auditory modality. We have performed a series of four studies in which we compared auditory and visual presentation modalities in the CTP. The first study varied modalities during probe or irrelevant presentation, but the targets and nontargets were always presented visually. Detection accuracies were superior in the visual probe or irrelevant presentation modality (Deng, Rosenfeld, Ward, & Labkovsky, 2016; Rosenfeld, Ward, Frigo, Drapekin, & Labkovsky, 2015). The Deng et al. (2016) study also systematically manipulated target and nontarget modalities in addition to probe and irrelevant modality, using either solely auditory target or nontarget presentation in the first experiment or combined visual and auditory presentation of targets and nontargets in the second. It was found that modalities of the probe-irrelevant and target/nontarget presentations interact. The most effective combination of presentation modalities always involved visual presentation of probes and irrelevants and either solely visual presentation of targets/nontargets, or combined visual and auditory presentation of targets and nontargets.

8 | CONCLUSIONS AND LIMITATIONS

The P300-based CTP has proven resistant to various countertactics and countermeasures while retaining good accuracy in situations where the probe is known in advance. The protocol is not affected in the laboratory by motivational manipulations and is most accurate when probe and irrelevant stimuli are presented visually. There are, however, some present limitations on the possible future application of the P300 CTP for field use: (a) As is true of all information detection protocols, in many crime situations it is often difficult to generate a sufficient number of informational items so as to compose an adequate test. Indeed, it takes a great deal of effort to assemble a scientifically valid CIT, as pointed out in the last chapter of Lykken (1998). Some law enforcement agencies may resist generating such efforts, in addition to resisting the effort required to learn electrophysiological technique. (b) As discussed earlier, studies have demonstrated an effective countermeasure against the "proactive" version of the CTP that is needed to prevent criminal acts in antiterror and other scenarios in which investigators do not have advance knowledge of specific probe stimuli (Lukács et al., 2016). This is an issue for future research. (c) There is resistance in the legal community to the use of psychophysiological procedures that purport to identify deception (Rosenfeld et al., 2013). It is argued that determination of who is lying is the job of the jury. Of course, P300-based and other CITs do not claim to detect deception; the claim usually made is limited to identification of recognized information, from which one may sometimes infer deception. This matter has been discussed, and the author urged that psychophysiologists eager to see courts adopt physiologically based CITs make efforts to present to the legal community P300 and other physiological evidence of recognition simply as novel and well-validated types of forensic data, certainly as worthy of consideration by juries as fingerprint, blood spatter, and DNA evidence (Meixner, 2012).

REFERENCES

Abootalebi, V., Moradi, M. H., & Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *International Journal of Psychophysiology*, *62*, 309–320. https://doi.org/10.1016/j.ijpsycho.2006.05.009

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

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369. https://doi.org/10.1038/35066572

Anderson, M. C., & Levy, B. J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science*, *18*(4), 189–194. https://doi.org/10.1111/j.1467-8721.2009.01634.x

Bergstrom, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological. Psychology*, *94*, 1–11. https://doi.org/10.1016/j.biopsycho.2013.04.012

Cutmore, T. R. H., Djakovic, T., Kebell, M. R., & Shum, D. H. K. (2009). An object cue is more effective than a word in ERP-based detection of deception. *International Journal of Psychophysiology*, *71*, 185–192. https://doi.org/10.1016/j.ijpsycho.2008.08.003

Deng, X., Rosenfeld, J. P., Ward, A., & Labkovsky, E. (2016). Superiority of visual (verbal) vs. auditory test presentation modality in a P300-based CIT: The complex-trial protocol for concealed autobiographical memory detection. *International Journal of Psychophysiology*, *105*, 26–34. https://doi.org/10.1016/j.ijpsycho.2016.04.010

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

Duncan-Johnson, C. C., & Donchin, E. (1979). The time constant in P300 recording. *Psychophysiology*, *16*, 53–55. https://doi.org/10.1111/j.1469-8986.1979.tb01440.x

Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Taylor & Francis. https://doi.org/10.2307/2532810.

Ellwanger, J., Rosenfeld, J. P., Hankin, B. L., & Sweet, J. J. (1999). P300 as an index of recognition in a standard and difficult match-to-sample test: A model of amnesia in normal adults. *Clinical Neuropsychologist*, *13*, 101–109. https://doi.org/10.1076/clin.13.1.100.1978

Ellwanger, J. E., Rosenfeld, J. P., & Sweet, J. J. (1997). P300 event-related brain potential as an index of recognition response to autobiographical and recently learned information in closed-head injury patients. *Clinical Neuropsychologist*, *11*, 428–432. https://doi.org/10.1080/13854049708400473

Ellwanger, J., Rosenfeld, J. P., Sweet, J. J., & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology*, *23*(1), 9–23. https://doi.org/10.1016/0167-8760(96)00035-9

Ellwanger, J. W., Tenhula, W., Rosenfeld, J. P., & Sweet, J. J. (2000). Identifying simulators of cognitive deficit through combined use of neuropsychological test performance and event-related potentials. *Journal of Clinical and Experimental Neuropsychology*, *21*, 866–879. https://doi.org/10.1076/jcen.21.6.866.850

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related potentials. *Psychophysiology*, *28*, 531–547. https://doi.org/10.1111/j.1469-8986.1991.tb01990.x

Guilmette, T. J., Hart, K. J., Guiliano, A. J., & Leinenger, B. E. (1994). Detecting simulated memory impairment: Comparison of the Rey fifteen item test and the Hiscock forced-choice procedure. *Clinical Neuropsychologist*, *8*, 238–294. https://doi.org/10.1080/13854049408404135

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical* and Experimental Neuropsychology, 11, 967–974. https://doi.org/10.1080/01688638908400949

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science*, *26*(7), 1098–1106. https://doi.org/10.1177/0956797615575734

Johnson, M. M., & Rosenfeld, J. P. (1992). A new ERP-based deception detector analog II: Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, *12*, 289–306. https://doi.org/10.1016/0167-8760(92)90067-l

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

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, *16*(2), 177–216. https://doi.org/10.1016/0010-0285(84)90007-0

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2015). Orienting versus inhibition in the Concealed Information Test: Different cognitive mechanisms drive different physiological measures. *Psychophysiology*, *53*, 579–590. https://doi.org/10.1111/psyp.12583

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the Concealed Information Test. *Psychophysiology*, *54*(4), 628–639. https://doi.org/10.1111/psyp.12825

Labkovsky, E., & Rosenfeld, J. P. (2012). The P300-based, complex-trial protocol for concealed information detection resists any number of sequential countermeasures against up to five irrelevant stimuli. *Applied Psychophsiology*, *37*, 1–10. https://doi.org/10.1007/s10484-011-9171-0

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. https://doi.org/10.1111/j.1469-8986.2007.00566.x

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex-trial protocol version of the P300-based Concealed Information Test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, *110*, 56–65. https://doi.org/10.1016/j.ijpsycho.2016.10.010

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

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

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*(9), 879–904. https://doi.org/10.1111/psyp.12239

Meijer, E. H., Smulders, F. T., Merckelbach, H. L., & Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology*, *66*, 231–237. https://doi.org/10.1016/j.ijpsycho.2007.08.001

Meixner, J., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based Concealed Information Test. *Psychophysiology*, *48*, 149–154. https://doi.org/10.1111/j.1469-8986.2010.01050.x.

Meixner, J. B. (2012). Liar, liar, jury's the trier? The future of neuroscience-based credibility assessment and the court. *Northwestern University Law Review*, *106*(3), 1451–1488

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Rosenfeld, J. P. (2005). "Brain fingerprinting:" A critical analysis. *Scientific Review of Mental Health Practice*, *4*, 20–37.

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). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.005

Rosenfeld, J. P. (Ed.). (2018). *Detecting concealed information and deception: Recent developments*. Cambridge, MA: Academic Press/Elsevier. https://doi.org/10.1016/B978-0-12-812729-2.00006-9

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based control-question lie detector analog: Algorithms for discriminating effects within individual waveforms. *Psychophysiology*, *28*, 320–336. https://doi.org/10.1111/j.1469-8986.1991.tb02202.x

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. https://doi.org/10.1016/j.ijpsycho.2005.06.002

Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *42*(1–2), 157–161. https://doi.org/10.3109/00207458808985770

Rosenfeld, J. P., & Donchin, E. (2015). Resampling (bootstrapping) the mean: A definite do. *Psychophysiology*, *52*(7), 969–972. https://doi.org/10.1111/psyp.12421

Rosenfeld, J. P., Ellwanger, J. W., Nolan, K., Wu, S., Bermann, R. G., & Sweet, J. (1999). P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model. *International Journal of Psychophysiology*, *33*(1), 3–19. https://doi.org/10.1016/S0167-8760(99)00021-5

Rosenfeld, J. P., Ellwanger, J., & Sweet, J. (1995). Detecting simulated amnesia with event-related brain potentials. *International Journal of Psychophysiology*, *19*, 1–11. https://doi.org/10.1016/0167-8760(94)00057-L

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. (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. https://doi.org/10.1016/j.ijpsycho.2013.08.012

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. https://doi.org/10.1111/j.1469-8986.2010.01024.x

Rosenfeld, J. P., Labkovsky, E., Davydova, E., Ward, A., & Rosenfeld, L. (2017). Financial incentive does not affect P300 (in response to certain episodic and semantic probe stimuli) in the complex-trial protocol (CTP) version of the Concealed Information Test (CIT) in detection of malingering. *Psychophysiology*, *54*(5), 764–772. https://doi.org/10.1111/psyp.12835

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. https://doi.org/10.1111/j.1469-8986.2008.00708.x

Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: A new method of lie detection. *International Journal of Neuroscience*, *34*(1–2), 125–129. https://doi.org/10.3109/00207458708985947

Rosenfeld, J. P., Reinhart, A. M., Bhatt, M., Ellwanger, J., Gora, K., Sekera, M., & Sweet, J. (1998). P300 correlates of simulated malingered amnesia in a matching-to-sample task: Topographic analyses of deception versus truthtelling responses. *International Journal of Psychophysiology*, *28*(3), 233–247. https://doi.org/10.1016/S0167-8760(97)00084-6

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

Rosenfeld, J. P., Sitar, E., Wasserman, J. D., & Ward, A. C. (2018). Moderate financial incentive does not appear to influence the P300 Concealed Information Test (CIT) effect in the complex-trial protocol (CTP) version of the CIT in a forensic scenario, while affecting P300 peak latencies and behavior. *International Journal of Psychophysiology*, *125*, 42–49. https://doi.org/10.1016/j.ijpsycho.2018.02.006

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205–219. https://doi.org/10.1111/j.1469-8986.2004.00158.x

Rosenfeld, J. P., Sweet, J. J., Chuang, J., Ellwanger, J., & Song, L. (1996). Detection of simulated malingering using forced choice recognition enhanced with event-related potential recording. *Clinical Neuropsychologist*, *10*(2), 163–179. https://doi.org/10.1080/13854049608406678

Rosenfeld, J. P., Ward, A. C., Drapekin, J., Labkovsky, E., & Tullman, S. (2017). Instructions to suppress semantic memory enhances or has no effect on P300 in a Concealed Information Test (CIT). *International Journal of Psychophysiology*, *113*, 29–39. https://doi.org/10.1016/j.ijpsycho.2017.01.001

Rosenfeld, J. P., Ward, A. C., Frigo, V., Drapekin, J., & Labkovsky, E. (2015). Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, complex-trial protocol for concealed autobiographical memory detection. *International Journal of Psychophysiology*, *96*, 16–22. https://doi.org/10.1016/j.ijpsycho.2015.02.026

Rosenfeld, J. P., Ward, A., Meijer, E. H., & Yukhnenko, D. (2017). Bootstrapping the P300 in diagnostic psychophysiology: How many iterations are needed? *Psychophysiology*, *54*(3), 366–373. https://doi.org/10.1111/psyp.12789

Rosenfeld, J. P., Ward, A. C., Wasserman, J. D., Sitar, E., Davydova, E., & Labkovsky, E. (2018). Effects of motivational manipulations on the P300-based complex-trial protocol for concealed information detection. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception: Recent developments* (pp. 125–144). London, UK: Academic Press/Elsevier. https://doi.org/10.1016/B978-0-12-812729-2.00006-9.

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). 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. https://doi.org/10.1016/S0167-8760(00)00154-9

Sweet, J. J. (1999). Malingering. *Forensic neuropsychology: Fundamentals and practice* (pp. 255–285). Lisse, Netherlands: Swets & Zeitlinger.

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Van Hooff, J. C., Brunia, C. H. M., & Allen, J. J. B. (1996). Event-related potentials as indirect measures of recognition memory. *International Journal of Psychophysiology*, *21*, 15–31. https://doi.org/10.1016/0167-8760(95)00043-7

Van Hooff, J. C., & Golden, S. (2002). Validation of an event-related potential memory assessment procedure: Intentional learning as opposed to simple repletion. *International Journal of Psychophysiology*, *16*, 12–22. https://doi.org/10.1027//0269-8803.16.1.12

Van Hooff, J. C., Sargeant, E., Foster, J. K., & Schmand, B. A. (2009). Identifying deliberate attempts to fake memory impairment through the combined use of reaction time and event-related potential measures. *International Journal of Psychophysiology*, *73*(3), 246–256. https://doi.org/10.1016/j.ijpsycho.2009.04.002

Ward, A. C., & Rosenfeld, J. P. (2017). Attempts to suppress episodic memories fail but do produce demand: Evidence from the P300-based complex-trial protocol and an implicit memory test. *Applied Psychophysiology and Biofeedback*, *42*(1), 13–26. https://doi.org/10.1007/s10484-016-9348-7

Winograd, M., & Rosenfeld, J. P. (2011). Mock crime application of the complex-trial protocol (CTP) P300-based Concealed Information Test. *Psychophysiology*, *48*, 155–161. https://doi.org/10.1111/j.1469-8986.2010.01054.x

How to cite this article: Rosenfeld JP. P300 in detecting concealed information and deception: A review. *Psychophysiology*. 2019;e13362. https://doi.org/10.1111/psyp.13362