P300 in detecting concealed information and deception.

J. Peter Rosenfeld, Department of Psychology, Northwestern University.

For *Psychophysiology*, special issue on P300, ed. by John Polich

---

Introduction: At the 1983 meeting of the Society of Psychophysiological Research, Fabiani et al. presented a study later published as Fabiani et al., (1986). This study reported that if a group of subjects learned a list of say 50 words on Day 1, then, two days later were tested on a larger list of, say, 200 words consisting of the 50 learned words embedded randomly in the larger list containing 150 novel words, a P300 ERP appeared in the average of ERP responses to the previously learned words which was not present in the average of the ERP responses to the novel words. Based on this result, it occurred to Rosenfeld et al., (1987, 1988) that the protocol used in that study could be easily adapted for use in a guilty knowledge or concealed information test (CIT).

Early studies of the P300 CIT: The classic CIT was invented by Lykken in 1959, and did not utilize ERPs as the dependent index of recognition, but measured the skin conductance response (SCR), as a sign of autonomic (ANS) arousal, probably related to the orienting reflex (see klein Selle et al., 2017). Although Lykken's ANS-based CIT was reasonably accurate (about 90% correct classification), one could argue that the ANS was only indirectly responsive to specific stimuli, and that the P300 ERP, increasingly linked to memory processes by the Donchin lab (in the 1980s), might be more accurate as a dependent index of recognition. Thus in 1987 and 1988, the first full-length studies of the P300-based CIT were published (Rosenfeld et al., 1987, 1988). In these studies, each

student subject selected a small item in a black box about 3 by 8 by 10 inches. There were 10 quasi-valuable items in that box—a 10-dollar bill, a wrist watch, a portable radio, etc. After the selection was made, and the selected items secreted in the pocket or purse, the subjects had EEG electrodes attached to their scalps. It was found that in nine of 10 subjects, a P300 appeared in response to the chosen item but not to the other items. The reported 90% accuracy was about the same as that reported by Lykken, (1959), although in subsequent studies, the autonomic CIT typically reported 80-90% accuracy, whereas the P300 CIT reported slightly higher accuracies—85%-100%.

The Donchin lab was also aware of the potential of P300 as a sign of recognition of "guilty knowledge" (crime-relevant information), because in 1991, they published a P300-based CIT study, based mostly on a mock espionage scenario (Farwell & Donchin, 1991), part of which had been reported earlier at the 1986 meeting of the Society of Psychophysiological Research. They reported a study of four previously admitted wrongdoers on the University of Illinois campus. Details of their actual crimes were detected with P300. A problem with this study was, however (as noted in Rosenfeld, 2005, 2012), the question of whether the P300 indexed their recognition of crime details when recognized during the P300 test, as opposed to the recognition of well-rehearsed facts admitted after the fact during multiple interrogations by parents, administrators, and campus security officials.The same issue could have been raised about the main study reported in the 1991 paper, in which student participants committed mock espionage acts whose details were rehearsed with the experimenters to a standard of perfection prior to the P300 test.

A few months prior to the Farwell & Donchin, (1991) paper, Rosenfeld et al., (1991) reported a study in which student participants were probed with P300 about anti-social/illegal acts from their pasts (cheating on tests, plagiarizing papers, using false IDs, and the like). This testing on more semantic information did not require rehearsal, however experimenters did ask subjects prior to the testing to examine a list of acts, and check the ones which applied to them. They did this only once, and they retained their lists (although experimenters recorded their answers with a hidden video camera), nevertheless, one could still ask whether or not the P300s in response to their guilty acts would have occurred had experimenters not used the checklists. Johnson & Rosenfeld, (1992) remedied this potential confound by testing (with a P300 CIT) subjects on only one item, designated *a priori*, without using any pre-test checklist. This was an item that probed about cheating on tests: The Rosenfeld lab had learned from the 1991 study that about half the subjects from the same population in the earlier study privately acknowledged cheating on tests, so it was assumed (correctly, as it turned out) that running a new sample from this same population would provide guilty and innocent groups of approximately equal sizes. Experimenters verified "ground truth" by giving the checklists *after* the P300 CIT in conditions of perceived privacy, as in Rosenfeld et al., (1991). It is noted that in Farwell & Donchin, (1991), Rosenfeld et al., (1991), and Johnson & Rosenfeld, (1992), the overall diagnostic accuracy was about 85 to 90%. How was guilt diagnosed individually?

<u>Diagnostic and definitional Issues:</u> Donchin's lab, which has introduced into the field so many novel methodological developments over the years, came

up with the idea of applying bootstrapping (Efron & Tibshirani, 1994) to the issue of individual diagnosis based on P300.

To fully appreciate this, it should first be understood that in the early P300 CITs (Farwell & Donchin, 1991; Rosenfeld et al., 1988, Rosenfeld et al., 1991), there were three kinds of critical stimuli used; the Probe (P), Irrelevant (I) and Target (T). Thus the early protocol was named the "3-stimulus protocol", 3ST, in Rosenfeld et al., (2006). In both labs, probes were items that were directly relevant to the information sought. They would be the specific murder weapon (e.g., 356 magnum) in a crime. Other items from the same category (guns) were presented as irrelevants (e.g., .38 colt, .45 automatic, .22 Beretta, etc.). Probes were infrequently presented (e.g., about 15% of trials) whereas irrelevants had a 70 % probability. The remaining 15% of the probability space was for target stimuli; these were irrelevant stimuli but pre-defined as requiring a unique response: Subjects were typically instructed to press a right mouse button to recognized targets (e.g. .32 colt) but the left mouse button to all other items (Ps and non-target Is.). Since both Ps and Ts were rare and meaningful for guilty Ss (rareness and meaningfulness being the antecedent conditions for P300 elicitation; Johnson, 1986), both these stimulus types were expected to elicit P300s. The frequent, non-target irrelevants were not expected to elicit P300, or if they did, it would be a much smaller P300.

Over the years, there have been two basic forms of the 3ST: 1) the single probe protocol as used by Rosenfeld et al., (1988,1991,1995), in which in each block of trials, there is 1 P, 1 T, and 4-7 different Is, each repeated about 20-40 times; and 2) the multiple probe protocol (utilized in Farwell & Donchin, 1991)

consisting of several differing P, I, and T stimuli, in the similar ratio of 1:4:1, each item repeated less often. Rosenfeld et al. (2007) found that the multiple probe protocol is more demanding than the single probe version, as indexed by RTs, and that consequently, the single probe protocol is more accurate.

In the Rosenfeld lab, the approach to individual diagnosis involves the classic "CIT effect" (klein Selle , Verschuere, & Ben Shakhar, 2017), which is the difference between the physiological indices in response to probe versus irrelevant stimulus (P-I P300 amplitude difference in the P300 CIT). In this situation, the T stimulus is used only as an attention holder: The subject must attend to all the stimuli in order to not miss the targets, responses to which constitute his explicit task. If (s)he misses targets, (s)he is not processing the stimuli, and the results should not be retained. The bootstrap technique (see Rosenfeld & Donchin, 2015, Rosenfeld et al., 2016) involves repeated resampling *with replacement* (from an subject's set of single sweep P and I samples) of P-I P300 amplitude differences 100-1000 times, with averaging of each set of iterated P300s. If 90% of these iterated mean differences are positive (P>I) then the subject is diagnosed as knowledgeable of the probe, from which guilt may be inferred. For the innocent or unknowledgeable subject, the probe is not known, so is, in effect, simply another irrelevant stimulus, and the expected value of P>I differences over many iterations as aboveis 50%. This method was called the bootstrapped amplitude difference method or "BAD" (Soskins et al., 2001).

In the Donchin lab (Farwell & Donchin, 1991) a more elaborate bootstrap computation was performed: It was assumed that the task relevant and rare

target stimulus should evoke P300, and so should the probe in knowledgeable subjects. Thus the P-T cross-correlation across most of the epoch will be high. In contrast, in the unknowledgeable subject, the probe is just another irrelevant so that the P-I correlation will be high. Thus, the probe, target, and irrelevant single sweeps were each re-sampled 100 times and the P-I correlations and P-T correlations over trials were obtained. If 90% of the comparisons show the P-T correlation > P-I correlation, the subject is determined to be knowledgeable. On the other hand, if 90% of the iterated comparisons show P-I > P-T correlation, then the subject is declared unknowledgeable. (Subjects whose comparisons do not yield either knowledgeable or unknowledgeable decisions are called indeterminate.) This method was called the bootstrapped correlation analysis of disparity or “BC-AD” (Soskins et al., 2001).

John Allen also did some early studies of P300 in the CIT; e.g., Allen, Iacono, & Danielson (1992). In these studies, a highly original Bayesian analysis was applied with great success to the issue of individual diagnosis. Abootalebi et al., (2006) utilized a wavelet classifier method addressed to the diagnostic challenge. Other methods have been introduced over the years, but as discussed in Rosenfeld (2011), there has never been a convincing comparison of the various methods, because the extant comparative studies then as now were never systematic, that is, they never compared studies all using similar threshold criteria, nor definitions of P300 measurement, e.g., base-to-peak vs. peak-to-peak, and so on.

Regarding P300 measurement in the P300-based CIT, extending Duncan-Johnson & Donchin (1979), Soskins et al. (2001) studied the (pre-stimulus)

baseline-to-peak (b-p) P300 (at Pz) recording at a high pass filter setting of .01 Hz *versus* the recording one sees at a more typical filter setting of .3Hz. At the former setting, one sees the superimposed P300s in the upper half of Fig. 1, below. The .3 Hz filtered recording is seen in the lower half of the figure. The down pointing arrows indicate the P300 peak in each case. The capacitive coupling of the lower recording causes the recovery slope of P300 to return to baseline and overshoot it into an apparent negative component, ("NEG") indicated in the lower wave pair with an up-pointing arrow. Soskins et al. (2001) found that the amplitude of NEG strongly correlated positively with the P300 recovery slope from the .01 Hz recording (Fig. 1 top). P300 amplitude and its recovery slope should be in principle, orthogonal, since the recovery slopes can vary over trials having the same peak amplitude. It was thus hypothesized that the recovery slope (or its correlate, the NEG wave) may thus provide information orthogonal and in addition to the peak amplitude of P300. On this basis Rosenfeld and colleagues have always advocated the p-p recording at .1 to.3Hz filter settings as the better P300 index in the P300-based CIT than any b-p recording, and have found that the p-p index is 25% -35% more accurate than any b-p recording (Rosenfeld, 2011). This finding has been replicated independently by Meijer et al., (2007), Cutmore et al. (2009), and Lukacs et al., (2016). Rosenfeld (2011) does not advocate use of the p-p measure for use in theoretical studies with P300, since the p-p recording is not a pure P300 peak amplitude measure.

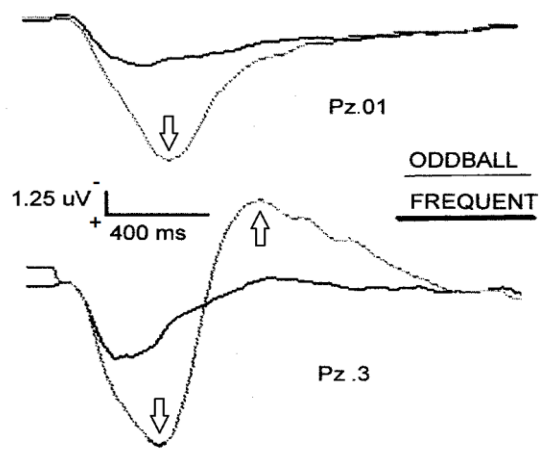

Fig. 1 Adapted from Soskins et al. (2001), a recording of P300s to frequent (p=.8) and rare/oddball (p = .2) verbal stimuli, recorded at low pass filter settings of .01 (top) and .3 (bottom), both from the same Pz location.

Some early applications of P300: **As noted, the early P300-based CITs were conceptualized as having relevance for forensic scenarios (Rosenfeld et al., 1988), employee screening (Rosenfeld et al., 1991, Johnson & Rosenfeld 1992, Mertens & Allen, 2008), and in espionage scenarios, (Farwell & Donchin, 1991). However, neuropsychologist colleagues alerted the field that the ability to detect denied knowledge (with P300) was also of potential use in the detection of *malingered cognitive deficit*. The first reports of this application were by Rosenfeld et al., (1995) and Ellwanger et al. (1996). These workers tested simulated malingers—university students instructed to malinger—on denied autobiographical knowledge such as birthdates, phone numbers, mothers' maiden names, as these item types appeared as stimuli in an autobiographical oddball paradigm. Although as instructed, subjects behaviorally denied recognition of these self-referring items, they were all potent elicitors of P300.**

It may have been noted, with respect to this approach, that patients with true head injury have been shown to have smaller than normal P300s, and thus,

suspected malingerers from this population might lack the requisite P300 ERP required by such a P300-based test. Ellwanger et al. (1997) pursued this issue and confirmed that although P300s were reduced in head injury patients, there was nevertheless a clear difference in their P300s in response to recognized versus unrecognized information.

Van Hooff and colleagues (Van Hooff et al., 1996, 2002) followed up this approach by applying the 3ST to a word recognition task similar to that of Fabiani et al., (1983). It was seen that recognized (previously learned) non-target words evoked P300. Since there was an interval of one day between learning and testing, the authors suggested that their procedure might be useful when "the integrity of memory is in question" as in cases of malingered cognitive deficit. This approach culminated in van Hooff et al. (2009), in which there was a normal group whose members were performing to the best of their abilities, and a sample from the same population, but asked to feign a credible memory disorder. ERPs, as well as RT, and performance on the Amsterdam Short-term Memory test were recorded. Although memory test results and RT data clearly distinguished the two groups, P300 amplitude and P300 scalp distribution did not. It was concluded that the discrepancy between behavioral and ERP data could be interpreted "as evidence of intentional underperformance," a most useful index to support a claim of malingering.

Rosenfeld et al. (1996,1998,1999) and Ellwanger et al. (1999, 2000) appreciated that there were limits to the use of the autobiographical oddball paradigm in diagnosis of malingered cognitive deficit. For one thing, most malingerers are not so unsophisticated as to verbally state that they don't

recall, for example, their birthdate, when in fact they may have just filled out an information card in which they provided that information in a neighboring office. The *behavioral* "Multidimensional Memory Test" (MDMT, see Guilmette et al., 1994; Hiscock & Hiscock, 1989) was developed as an entrapment procedure to catch these people. It is a simple matching-to-sample test: A sample 3-digit number is presented followed 5-10 s later either by a match or mismatch, each of which has a 50% chance of appearing. It is reported that that normal persons and even patients with moderate head injury perform well on this test, unless they are involved in litigation. A 90% correct criterion was proposed as a cutoff for a diagnosis of malingering (Guilmette et al., 1994), however Ellwanger et al. (1997, 1999) reported that in a population of 15 (non-litigating) closed head injury patients, two of these people scored < 80% correct on the behavioral MDMT; i.e., this behavioral test was far from infallible.

Thus, Rosenfeld et al. (1996,1998,1999) and Ellwanger et al. (1999, 2000) modified and enhanced the MDMT to include ERP data, P300 in particular. Most importantly, the relative probabilities of match and mismatch were changed to 17% and 83% respectively, and the interval between the sample number and the subsequent test number was reduced to 3-5 seconds. It was found that the P300s and P300 scalp amplitude distributions of simulated malingerers could be readily discriminated from those of control subjects, however at best, the accuracy of classification never exceeded 75%. This figure applied also to later papers in which Ellwanger introduced yet another modification of the MDMT in which, following the presentation of the sample number, a series of six numbers followed including one match randomly placed in a series of mismatches. However Ellwanger et al. (1999), further altered the original P300-enhanced

MDMT such that match and mismatch probabilities were 11% and 89% respectively. In that study, classification accuracies of malingering simulators and non-simulators were .8 (sensitivity) and 1.0 (specificity) respectively for an overall 90% accuracy.

Another recently published application for P300 as a sign of recognition in the 3ST was by LeFebvre et al. (2007). This work advanced the field by applying P300 as a sign of recognition to pictorial, specifically facial identification, in the context of eyewitness testimony accuracy. LeFebvre and colleagues had subjects watch crimes enacted on videos in which a culprit entered the premises of a victim and stole a laptop computer. In the subsequent 3ST 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 subject –who was the "witness"—were irrelevants. The whole stimulus set was regarded as a lineup analog. Subjects were tested either immediately (one minute) after watching the video, one hour later, or one week later. There was one other condition in which the actual culprit was absent from the "lineup." The individual diagnostic procedure involved comparison of the P300 averages, converted to z-scores, for each stimulus versus the averages to all stimuli from each of 6 parietal and centro-parietal scalp sites. A *correct identification* occurred when the z-score from the actual probe exceeded that of the grand average by 2 or more. A *misidentification* occurred when one of the irrelevant faces evoked a P300 whose z score = or was > 2. If none of the P300 z scores reached 2, then the decision was either *false rejection* if there actually had been a probe present in the "lineup", or a *correct rejection* if the lineup lacked a culprit.

The major results were that in the immediate and one hour delay conditions, 79% and 83% accuracies were obtained, however accuracy fell to 58% with a one week delay. Moreover, 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."

<u>The Countermeasure Issue in the 3ST and the new Complex Trial Protocol:</u> Rosenfeld et al. (2004) appreciated that if in the 3ST, the experimenter identified an irrelevant stimulus as a target that required a unique button press, and that this stimulus led to a P300 response, then a subject could also decide secretly for himself to make secret responses to other irrelevants, which would lead to other P300s. If this was done, then there would be no differences in the averages of probe versus irrelevant responses, thus destroying the diagnostic basis of the 3ST, as demonstrated in Rosenfeld et al., (2004): There were two experiments in that paper. In the first, subjects 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: jewelry type, lining paper color, owner's name, operation name, furniture type, items on top of desk. (This structure was based on Farwell & Donchin, 1991.) There was 1) a simply guilty (SG) group, 2) a guilty group taught countermeasures (CM), and 3) an innocent (IN) group whose members stole nothing.

Subjects were taught six simple CMs for each category of stimulus, and told to execute one of these CMs for each irrelevant item that appeared within the same category. (The CMs included acts like non-noticeable right or left finger presses, toe wiggles, and mental images). Results in terms of diagnostic accuracies are given in the table below using our bootstrapped amplitude difference "BAD" method and the Farwell & Donchin (1991) bootstrapped correlation analysis of disparity "BC-AD" method, both using a 90% bootstrap criterion as described above:

Diagnoses of Guilty

Amplitude Difference (BAD) method,p=.1		
Guilty Group	Innocent Group	CM Group
9/11(82%)	1/11(9%)	2/11(18%)
Cross-Correlation (BC-AD) Method, p=.1		
6/11(54%)	0/11(0%)	6/11(54%)

Table 1: Results of Experiment 1 in Rosenfeld et al. (1991)

It is clear that with the BAD method, CMs reduce accuracies from 82% to 18%. With the BC-AD method, accuracies are poor in the first place, and just as bad with CMs. (The six subjects detected in the guilty group are not the same as the six detected in the CM group, suggesting random detection.)

In the second experiment in Rosenfeld et al., (2004), CMs were taught to subjects simulating malingered amnesia for their own birthdates. We used a one-probe protocol in which the stimuli were each subject's birthdate, four

irrelevant dates, and one designated target irrelevant date, each date repeated 30 times. The subjects were run in three successive weeks, in the first as simply guilty, in the second with CMs, and in the third with instructions to repeat the first week. The CMs were as before: for each of the four irrelevant dates, make either a left or right finger press, or toe wiggle. This is the one stimulus protocol of Rosenfeld et al., (1995). The results are in the next table (BAD and BC-AD as above.):

WEEK		BAD	BC-AD
1	no CM	12/13(.92)	9/13(.69)
2	CM	6/12(.50)	3/12(.25)
3	no CM	7/12(.58)	3/12(.25)

Table 2: Results of Experiment 2 in Rosenfeld et al. (1991)

It is evident that the CMs impact this protocol also, whether diagnoses are made with the BAD or BC-AD method. The CM effects persist into the third week when there were no CMs used, as confirmed with a return of RTs from a very elevated level in the second week, when CMs were used, to the same levels for all (P,T, and I) stimuli as seen in the first week. It was clear that a new CM-resistant protocol was needed.

Rosenfeld et al. (2008) reasoned that demand aspects of the 3ST made it vulnerable to CMs: Each trial contained a dual task: the explicit Target-Non-target discrimination and the implicit Probe-Irrelevant discrimination. That is, on each trial the subject needed to explicitly decide whether or not the target

response (right button press) was in order. However, we also assumed that when the probe or irrelevant (both non-targets) stimulus was presented, there would also be an *implicit* response conflict even though both P and I require the same left button press. Donchin, Kramer, and Wickens (1986) summarized many studies in noting 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 one sees when there is only the oddball task in effect. Rosenfeld et al. (2008) hypothesized that P300 to the probe in the 3ST is not as large as it could potentially be because of the division of attention between the two (explicit vs. implicit) tasks. They also devised a new P300 CIT in which the two tasks were separated, as shown in the following figure.

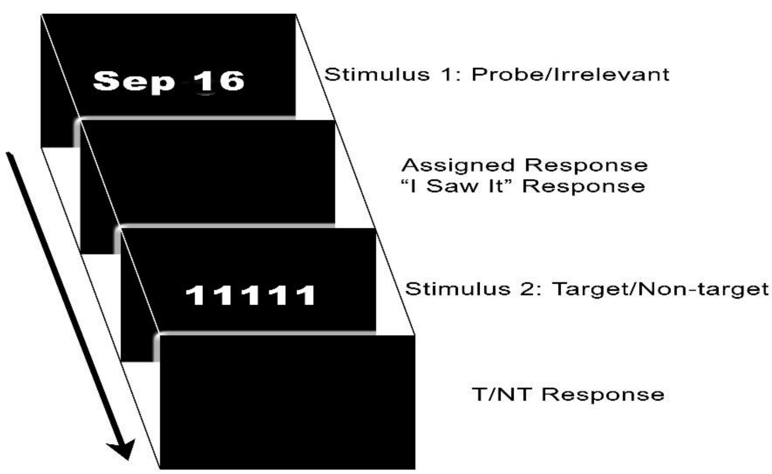

Fig. 2. The Complex Trial Protocol

It is seen in this figure that there are two stimuli presented on each trial, S1, either a probe or an irrelevant, and then about one to two seconds later, either a target or non-target. A date is shown in Fig. 1 above as S1 when one is

probing to see if a suspected malingerer can recognize the birthdate. S2 shows either a target number string (11111) or a non-target number string (22222, 33333,…to 55555). The response to S1 is the same left button press on the left hand mouse no matter which of S1 or S2 is presented. Subjects are warned in advance that there will be un-announced tests at random intervals (about every 10-40 trials) as to the identity of the S1 presented, and that there will be an adverse consequence if more than one of these pop quizzes is incorrectly answered; this forces attention. 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 the naming of this protocol as the *Complex Trial Protocol* (CTP).

In the first test of the CTP, birthday probes were used, as for the data of Table 2 above, and the design of the experiment was the same: 1) a week with no CMs, 2) a week with CMs, and finally 3) a week with no CMs again. The diagnostic accuracies based on the BAD test were unaffected by CM use, as follows:

WEEK	Hit Rate
Week 1 (no CM):	11/12
Week 2 (CM):	10/11
Week 3 (no CM):	11/12

Table 3: Results of Rosenfeld et al. (2008)

In this study, a unique CM was performed for each of the four irrelevants, however, in subsequent studies, we reported that CMs against any number (one to four) of irrelevants did not reduce the accuracy of the CTP below 90% (Rosenfeld & Labkovsky, 2010; Labkovsky & Rosenfeld, 2012) , including studies based on episodic memory information as in a mock crime scenario (Winograd & Rosenfeld, 2011).

A recent and timely application of the P300-based CTP to the anti-terror challenge was reported by Meixner & Rosenfeld (2011). In this scenario, the aim is to prevent the criminal act by detecting it in the planning stage. Also, in this scenario, one must arrest the suspects and probe their memories for planned but not yet committed acts. Meixner and Rosenfeld had simulating terrorists read informational brochures, then choose the dates, place and time of their mock acts of terror, and write a letter to mock terrorist chieftains summarizing and justifying their attack choices. A control group wrote a letter about vacation plans but took the same CTP with three separate blocks (testing on recommended date, method, and city of planned terror act) as the mock terrorists. The BAD method with 1000 bootstrap iterations was applied to each CTP block for both terror and control groups.

The number of bootstrap iterations where the known P was compared to the average Irrelevant averaged 966/1000 in terror Ss, vs. 546 (close to chance = 500) in controls. This yielded 100% accuracy with 12/12 correct guilt detections with 0/12 false positives for a perfect ROC area score of 1.0. But the point of detection here, as noted above, is to identify suspects *before* the act is committed. To simulate this scenario, it was assumed that the largest P300

within a guilty subject among average P300s to all stimuli (P and I) would be the average to the probe. This average was compared in each subject to the next largest average P300 for that subject, assuming it to represent the largest response to an irrelevant stimulus. That test, called the "blind Imax test" yielded 10/12 accurate detections in guilty Ss with no false positives, for an area under the ROC curve = .979, a value indexing excellent discriminability. (Comparing the maximum P300 to the average of all remaining P300 averages led to an unacceptable number of false positives.)

<u>Independent Replication of the CTP, but Novel Countermeasures:</u> Lukacs et al., (2016) completely replicated the main findings of Rosenfeld et al., (2008) and other related studies (summarized in Rosenfeld et al., 2013) relevant to scenarios in which the probe is known in advance by experimenters. They found that if the probe P300 average is compared in the bootstrap with the average of all irrelevants, the detection rate was >90% when the usual countermeasures were used (making irrelevants into secret targets with covert responses). However, Lukacs et al. developed a novel countermeasure which significantly reduced the sensitivity of the CTP when the "blind Imax" test was used, as in anti-terrorist proactive protocols (such as Meixner & Rosenfeld, 2011). The complex countermeasure was to have subjects a) secretly and mentally say a word like "dog" to themselves when seeing either the probe or any one of four irrelevant items b) secretly and mentally say a word like "cat" to themselves when seeing the remaining two irrelevant items used. Presumably, the last two irrelevants became the rare and meaningful items capable of evoking the largest P300s, thus defeating the blind Imax test. Our view (which we are presently

testing) is that this countermeasure is difficult to perform, and requires much practice.

Other counter-tactics have also been recently attempted against the CTP, in particular the technique of mental suppression (e.g., Anderson & Levy, 2009). Evidence for suppression is largely based on demonstrations of suppression of response term memories in paired associate learning (Anderson and Green, 2001). Further support has come from recent reports of ERP correlates of such suppression: Bergstrom et al. (2013) and Hu et al. (2015) both reported that the P300 sign of recognition could be voluntarily suppressed. The former report however used the 3ST, and as additionally pointed out by Rosenfeld et al. (2017a), the claim of P300 suppression was confounded by differences in amount of rehearsal between key comparison groups. Hu et al. (2015—from the Rosenfeld lab) used a CTP, but uniquely, used a 50-50 target/no-target ratio (versus the usual 20-80 ratio) which proved to increase task demand (probably reducing P300 amplitude; see Ward & Rosenfeld, 2017). Moreover, Hu et al. showed suppression only in the base-peak P300. *With the generally preferred peak-peak method, no suppression was seen.* Indeed, Rosenfeld et al. (2017a) and Ward & Rosenfeld (2017) unambiguously showed either no effect or *enhancing* effects of suppression instructions on P300 signs of episodic and semantic recognition in the CTP.

<u>Recent studies of motivational and other effects on the CTP.</u> In a systematic study (10 groups of 15-19 subjects each) of the effects of motivational (financial and instructional) manipulations on the P300 in the CTP, Rosenfeld et al. (2017b, 2018a, 2018b) demonstrated that manipulations which clearly produce dependent behavioral effects, do not affect the P300 sign of

recognition of episodic or semantic memories in both mock crime or simulated malingering scenarios. This was important because as recently reviewed by Meijer et al. (2014), the typically employed ANS response of SCR *is* responsive to motivational manipulations, meaning that what is obtained in the lower stakes lab situation is not necessarily applicable to the field situation. Fortunately, this appears not to be the case for P300.

In standard detection of deception procedures as presently used in law enforcement, the questions are posed to subjects by operators, i.e., in the auditory modality. Thus Rosenfeld et al. (2015) and Deng et al. (2016) performed a series of four studies in which they compared auditory and visual modalities in the CTP. The first study varied modalities during probe or irrelevant presentation, but the targets and non-targets were always presented visually. Detection accuracies were superior in the visual modality. Deng et al., (2016) also systematically manipulated target and non-target modalities in addition to probe and irrelevant modality, using either solely auditory target and non-target presentation in their first experiment or combined visual and auditory presentation of targets and non-targets in their Experiment 2. It was found that modalities of the probe-irrelevant and target/non-target presentations interact, and that the most effective combination of presentation modalities always involved visual presentation of probes and irrelevants and either solely visual presentation of targets/non-targets, or combined visual and auditory presentation of targets and non-targets.

<u>Conclusions:</u> The P300 based CTP has proven largely resistant (but not immune) to various counter tactics and countermeasures while retaining good accuracy in situations where the probe is known in advance. The protocol is not

affected by motivational manipulations, and is strongest when stimuli are presented visually.

References:

Abootalebi, V., Moradi, M.H. & Khalilzadeh, M.A. (2006) A comparison of methods for ERP assessment in a P300-based GKT. *Int. J. Psychophys.*, 62, 309-320

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.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*(6826), 366–369.

Anderson, M. C., & Levy, B. J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science, 18*(4), 189–194.

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. *Biol. Psychol*. 94, 1–11. http://dx.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. *Int. J. Psychophysiol.,* 71, 185-192.

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.

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

Duncan-Johnson, C.C., Donchin, E., (1979). The time constant in P300 recording. *Psychophysiology* 16, 53-55.

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

Ellwanger, J., Rosenfeld, J. P., Hankin, B.L., & Sweet, J.J. (1999).P300 asan index of recognition in a standard and difficult match-to-sample test: A model of amnesia in normal adults. The Clinical Neuropsychologist, 13,101-109.

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.

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. *The Clinical Neuropsychologist*, 11, 428-432.

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. *J. Clin. and Exp. Neuropsychology*, 21, 866-879.

Fabiani, M., Karis. D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology,* 23, 298-308.

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

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. *The Clinical Neuropsychologist, 8*, 238-294.

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.

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.

Johnson, M.M. and Rosenfeld, J.P. (1992) A new ERP-based deception detector analog II: Utilization of non-selective activation of relevant knowledge. *Int. J. Psychophysiology*, 12, 289-306.

Johnson, R.R. (1986) A Triarchic model of P300 amplitude. *Psychophysiology*, 23, 367-384

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.

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.

Lukacs, G. et al. (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. *Int. J. Psychophysiol.,* 110 (2016) 56–65

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

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.

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.

Meixner, J. & Rosenfeld, J.P. (2011). A Mock Terrorism Application of the P300-Based Concealed Information Test, *Psychophysiology.* 48, 149-154

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

Rosenfeld, J.P., et al. (1987) Late vertex positivity as a guilty knowledge indicator: A new method of lie detection. Int. J. Neurosc., 34, 1987, 125-129.

Rosenfeld, J.P. et al. (1988) A modified, event-related potential-based guilty knowledge test. *Int. J. Neurosc.*, 42, 157-161.

Rosenfeld, J.P., Angell, A., Johnson, M., and Qian, J. (1991) An ERP-based control-question lie detector analog: Algorithms for discriminating effects within individual waveforms. *Psychophysiology,* 28, 320-336.

Rosenfeld, J.P., Ellwanger, J., and Sweet, J. Detecting simulated amnesia with event-related brain potentials. *Int. J. Psychhophysiol.*, 19, 1995, 1-11.

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. *The Clinical Neuropsychologist*, *10*(2), 163-179.

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.

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.

Rosenfeld, J.P. and Ellwanger, J.W. (1999). Cognitive

psychophysiology indetection of malingered cognitive deficit. In J.J.Sweet (Ed.) *Forensic Neuropsychology,* Lisse (Switzerland): Swets &Zeitlinger, publishers, pp. 287-312.

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

Rosenfeld, J.P. (2005)"Brain Fingerprinting:" A Critical Analysis. *Scientific Review of Mental Health Practice*, 4, 20-37.

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. *Int. J. Psychophysiology*, 60, 251-259.

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

Rosenfeld, J.P., 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.

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.

Rosenfeld (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* . Cambridge University Press. 63-89.

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. *Int. J. Psychophysiol.* 90, 118–134.

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

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. *Int. J. Psychophysiol.* 96 (2015) 16–22

Rosenfeld, J. P., Ward, A., Meijer, E. H., & Yukhnenko, D. (2016). Bootstrapping the P300 in diagnostic psychophysiology: How many iterations are needed?. *Psychophysiology*. Version of Record online: 7 DEC 2016 | DOI: 10.1111/psyp.12789

Rosenfeld, J.P., Ward, A.C., Drapekin, J., Labkovsky, E., & Tullman, S. (2017a) 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.

Rosenfeld, J. P., Labkovsky, E., Davydova, E., Ward, A., & Rosenfeld, L. (2017b). 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*. https://doi.org/10.1111/psyp.12835.

Rosenfeld, J.P., Ward, .C., Wasserman, J.D., Sitar, E., Davydova, & Labkovsky (2018a). Effects of Motivational Manipulations on the P300-Based Complex Trial Protocol for Concealed Information Detection. In *Detecting concealed information and deception: Recent Developments*, J.P. Rosenfeld, Ed., Academic Press/Elsevier.

Rosenfeld, J.P. Sitar, E., Wasserman, J.D., & Ward, A.C. (2018b). 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, p.42-49. DOI: 10.1016/j.ijpsycho.2018.02.006

Soskins, M., Rosenfeld, J.P., & Niendam, T. (2001). Thecase for peak-to-peakmeasurement of P300 recorded at .3 hz high pass filter settings in detection of deception. *Int. J. Psychophysiology*, 40, 173-180.

Sweet, J. J. (1999). Malingering. *Forensic neuropsychology: Fundamentals and practice*, 255.

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.

Van Hooff, J. C., and 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.

Van Hooff, J. C., Brunia, C. H. M., and Allen, J. J. B. (1996). Event-related potentials as indirect measures of recognition memory. *International Journal of Psychophysiology*, 21, 15–31.

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*. DOI 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