Appl Psychophysiol Biofeedback (2012) 37:1–10
DOI 10.1007/s10484-011-9171-0

# The P300-Based, Complex Trial Protocol for Concealed Information Detection Resists Any Number of Sequential Countermeasures Against Up to Five Irrelevant Stimuli

**Elena Labkovsky · J. Peter Rosenfeld**

Published online: 1 October 2011


**Abstract** We previously tested the P300-based Complex Trial protocol for deception detection against 2 and 4 countered of 4 irrelevant stimuli. The protocol detected 90–100% of these subjects with <10% false positives. We have also shown that Reaction Time (RT) to the first trial stimulus is increased (group effect) with countermeasure use. We also reported a new P900 component associated with countermeasure use when 2 of 4 irrelevants are countered. In the present study we report data from 4 new groups and re-present for comparison previously collected data to have 7 groups: an innocent control, a guilty group not using countermeasures, and 5 guilty/countermeasure groups who counter from 1 to all 5 stimuli presented (4 irrelevants plus a probe). Subjects were detected at rates varying from 92 to 100% in the 6 guilty groups (n = 12 or 13 per group). There was 1 false positive in 13 innocent subjects. Additionally, 41 of 60 CM users were identified with RT as using countermeasures. P900 appeared in the 2 groups using 2 and 3 countermeasures.

**Keywords** ERP · P300 · Concealed information test · Deception · Complex trial protocol

## Introduction

For almost a century, and with renewed intensity since September 11, 2001, there have been enormous efforts expended by governments and universities to develop an accurate deception test based on sound scientific principles. Both polygraph protocols using the measurements of autonomic nervous system activity (the Comparison Question Test, CQT, and the Concealed Information Test, CIT, (also called the Guilty Knowledge Test or GKT) have been alternatively advocated and criticized, as recently summarized in a long report by the National Research Council of the National Academy of Sciences, (National Research Council 2003). The methods in this paper represent the CIT approach largely preferred by the scientific community (National Research Council 2003) that has largely rejected the other (CQT) approach. In the CIT, many multiple choice questions are put to the subject about crime details (e.g., "Which was the murder weapon, was it (a) the hunting knife? (b) the baseball bat? (c) the rifle? (d) the pistol?"). There is only one correct choice (called the probe), and when it is put to the guilty subject (who denies the correct choice), the largest physiological response is expected in comparison to responses to the other choices (called irrelevants).

Among the problems with polygraphy raised by the National Research Council report is its potential susceptibility to *countermeasures*. As stated by Honts et al. (1996, p. 84), ''Countermeasures are anything that an individual might do in an effort to defeat or distort a polygraph test.'' The National Research Council report went on to state that ''Countermeasures pose a serious threat to the performance of polygraph testing because all the physiological indicators measured by the polygraph can be altered by conscious efforts through cognitive or physical means'' (National Research Council 2003, p. 4). More specifically, countermeasures are effective against both the polygraphic CQT (Honts et al. 2001), as well as against the polygraphic CIT, (Honts et al. 1996; Elaad and Ben-Shakhar 1991; Ben-Shakhar and Dolev 1996).

E. Labkovsky · J. P. Rosenfeld (✉)
Department of Psychology, Northwestern University, Evanston, IL, USA
e-mail: jp-rosenfeld@northwestern.edu

It was hoped and indeed expected that when the P300 Event-Related EEG Potential (ERP) was introduced as the dependent index of recognition in a CIT (Farwell and Donchin 1991; Rosenfeld et al. 1988, 1991), the countermeasure issue would be resolved. The eminent inventor of the GKT version of the CIT, Lykken (1998, p. 293), suggested about countermeasures to P300 CITs: "Because such potentials are derived from brain signals that occur only a few 100 ms after the GKT alternatives are presented…it is unlikely that countermeasures could be used successfully to defeat a GKT derived from the recording of cerebral signals." (Ben-Shakhar and Elaad 2002 expressed a similar view.)

Unfortunately, Rosenfeld et al. (2004) and Mertens and Allen (2008) showed that the original form of the P300-based CIT was vulnerable to simple countermeasures: subjects simply learned to make secret responses (e.g., toe wiggles) to the *irrelevant* items, converting them into covert *relevant* items that evoked P300s indistinguishable from the probe P300s, thus defeating the test. This prompted our lab to develop a novel P300 protocol which by design has thus far resisted previously effective countermeasures (Rosenfeld et al 2004; Mertens and Allen 2008) in three new studies (see Rosenfeld et al. 2008; Rosenfeld and Labkovsky 2010; Winograd and Rosenfeld 2011).

Indeed the novel *complex trial protocol* (CTP) has so far been the *only* physiologically based deception testing protocol reported that is resistant to CMs, and additionally, provides a simple index—reaction time (RT)—of the *use* of CMs by subjects. Thus, the test has so far typically identified recognition of concealed information as well the attempt by guilty subjects to counter the protocol—which likely constitutes *additional* evidence of a subject's criminal complicity. Moreover, even in the rare cases we occasionally encounter in which a subject whom we instruct how best to defeat our CTP succeeds in not showing the enhanced P300 indicator of guilty knowledge recognition, his RT index may still give away his attempt at non-cooperation—useful additional information for enforcement officials.

In the P300-based CTP, subjects are presented in each trial with two sequential tasks. The first (and critical) one involves responding immediately to a first stimulus (S1), either a probe (crime-relevant or key) or irrelevant stimulus with a single button press acknowledging perception; (therefore this R1 response is called the "I Saw It" response). The meaningful and rare probe, but not the irrelevant stimulus elicits the P300 sign of relevant probe recognition in a knowledgeable subject. Thus, the key indicator of guilty knowledge recognition is a larger probe than irrelevant P300. RT for the "I saw it" response *to irrelevant stimuli being countered* will typically be elevated in subjects taught to counter some or all irrelevant stimuli, since subjects need to take some moments to recall which countermeasure (CM) to make, possibly after recalling which, if not all, irrelevants need countering. In the CTP, about a second after S1 is presented, an S2 (either a target—requiring a unique button response—or non-target) stimulus is presented to maintain the subject's attention, which is also controlled with unpredictable, occasional probe identification tests (see "Methods").

In Rosenfeld et al. (2008) we utilized on each block of trials one probe and four irrelevant stimuli, shuffled and presented in random order. Subjects in the CM group were told to make a different CM for *each* of the four irrelevant stimuli. We chose this approach because it seemed to us then to be the most likely method of beating our new protocol as it did in the original P300-based CIT in Rosenfeld et al. (2004). However, one reviewer of the 2008 paper had written, "the CM manipulation in the current study might have enhanced the salience of probes because these were the only stimuli … that did not require the selection and execution of a specific CM." In other words, the strategy of countering *all* irrelevants but not the probe stimulus might have worked against the aim of beating the test, since it may have enhanced the probe P300 by adding extra (task relevant) salience to that afforded by its being rare and meaningful as a personally relevant item. Indeed Meixner and Rosenfeld (2010) recently confirmed that hypothesis, suggesting that a better method of beating our putatively CM-resistant CTP would have been to counter fewer than all irrelevant items. As it is important to the field to have a CM resistant test, the present studies systematically explore the effects of countering from one to all (four irrelevants plus the probe) stimuli used in the present CTP. We are interested in effects on P300 amplitude, as well as on our RT index of CM use.

We hypothesize about probe P300 that it will be prominent in groups using one, four, and five CMs, but not in groups using 2 and 3CMs. This is because P300 amplitude is reduced by simultaneous ongoing, *independent* tasks (Donchin et al. 1986) and it would appear that when CMs are to be executed to middle (2CM, 3CM) fractions of irrelevants, the task of discriminating to-be-countered and other irrelevant stimuli is greatest, since these groups need to be the most vigilant about whether or not a CM is required.

Additionally, we recently observed (Rosenfeld and Labkovsky 2010) that when subjects counter two of four irrelevants in the CTP, an apparently novel ERP we called P900 appears in the Cz and Fz, but not Pz, recordings of ERP responses to probe stimuli and also, occasionally in ERP responses to irrelevant stimuli which are not countered. This component was not seen in probe Pz recordings, nor in any ERP responses derived from innocent subjects, nor from guilty subjects *not* doing CMs. We suggested that

P900 would be useful *additional* information that CMs were in use, thus supplementing RT for this purpose. This component was never seen previously, for example, in subjects countering all four irrelevants as in Rosenfeld et al. (2008). We hypothesized in Rosenfeld and Labkovsky (2010) that the P900 represented a signal to a subject that no further responses to the signal would be required after stimulus acknowledgement. Thus, a probe stimulus was such a signal, but a countered irrelevant, in contrast, signaled that a further response (a CM) would be required. An un-countered irrelevant was a signal like a probe, but did require further cognitive processing. Thus we also study here the effects of various numbers of CMs used on the appearance of P900. Based on the combined earlier findings, our empirical hypothesis is that we will see P900 to the probe in groups where only two and three of four utilized irrelevants are countered (as in Rosenfeld and Labkovsky 2010), but not in the group where all four irrelevants are countered as in Rosenfeld et al. (2008), nor in the innocent (IN) and simply guilty (SG; no CMs used) groups. Because we reason that the group countering only one of four irrelevants is superficially similar to the SG group (that counters zero of four irrelevants) and will thus show a relatively large probe P300, we do not anticipate a large P900 here either. Neither do we expect probe P900s in the four-CM group (as none were seen in Rosenfeld et al. 2008, in which all four irrelevants were countered), nor in a group in which all five stimuli (probe plus four irrelevants) are countered since there is no stimulus in this situation signaling a subject that no further responses will be required.

## Methods

### Subjects and Groups

Participants were 86 students (42 females) from the introductory psychology class pool (see Table 1). The subjects were assigned to seven groups: (1) A simple guilty (SG) group—where one of the five S1 stimuli was the subject's birth date (Probe); (2) An innocent (IN) group where all S1 stimuli were irrelevant to subjects, and (3) five Countermeasure groups: 1CM group—similar to SG but subjects perform mental countermeasure (CM) to 1 Irrelevant; (4) 2CM group with CMs to 2 Irrelevants; (5) 3CM group—CMs to 3 Irrelevants; (6) 4CM group—CMs to all 4 Irrelevants; and (7) 5CM group—CMs to all five S1 stimuli, including the Probe. The IN, SG and 2CM group data were previously reported (Rosenfeld and Labkovsky 2010), and are presented again here for comparison. For these previously run three groups, the subjects were randomly assigned. For the four new groups separately run for this study, subjects were likewise (separately) randomly assigned.

**Table 1** Individual hit rates, probe versus Iall, $p = .9$ confidence

| Group | Detections |
|---|---|
| SG | 13/13 (100%) |
| 1CM | 11/12 (92%) |
| 2CM | 12/12 (100%) |
| 3CM | 12/12 (100%) |
| 4CM | 11/12 (92%) |
| 5CM | 11/12 (92%) |
| IN | 1/13 (8%) |

### Procedures

In the SG group (N = 13, 5 females, Mean age = 19.3, SD = 1.65), subjects first saw in a trial either a probe or one of four irrelevants presented for a 300 ms duration in random order on a computer monitor 1 M in front of them. This was S1. If a subject stated before testing that an irrelevant was personally meaningful, it was replaced with a different, confirmed meaningless irrelevant. Then, after about a one second delay (randomly varying 800–1,200 ms), S2, a string of numbers, was presented.

Participants were asked to make two responses within a trial, one to each type of stimulus. In response to S1, each participant was asked to randomly press one of the 5 buttons on a 5-button response box with the *left* hand, immediately after seeing any date (Fig. 1). This stimulus acknowledgement is called the R1 or "I saw it" response.

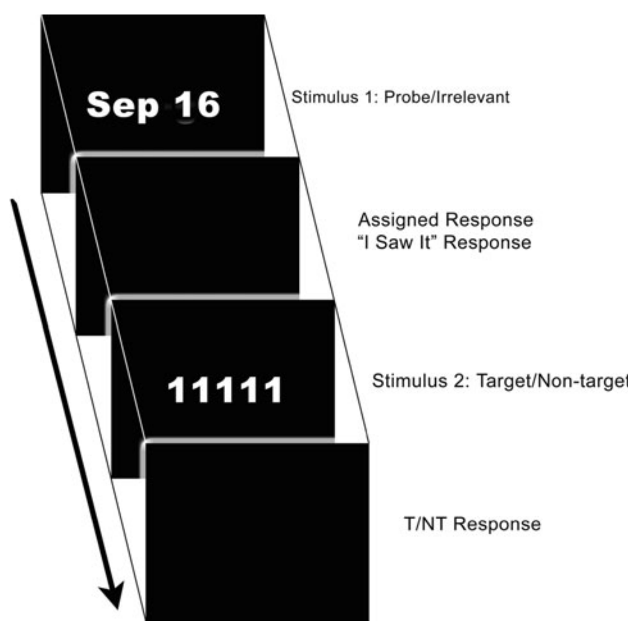


**Fig. 1** The events in the complex trial protocol using a date for S1 and a number string for S2

The subjects were instructed to do their best to press each of the five buttons approximately the same number of times as every other button. They were also told to try not to press the same button twice in a row, and not to develop a specific pattern of button presses. (Compliance was verified as in Rosenfeld and Labkovsky (2010) described below.) To further insure that the subjects were paying attention to the S1 stimuli (probes and irrelevants), the subjects were tested immediately after R1 at various unpredictable times throughout the experiment about which date they saw last. More than one failure would result in exclusion. There were no such exclusions here.

The random response “I saw it” task, as first described in Meixner et al. (2009), was used originally to focus attention on the first stimulus (S1) of the trial. We had noted in Rosenfeld et al. (2008) that CM-using subjects had larger P300s than simple guilty (no CMs) subjects. At that time we felt this was because we embedded the related (*non-independent)* CM task within the R1 (I saw it) task by requiring CM execution before R1, thus forcing extra attention to the S1 and enhancing its P300 (Donchin et al. 1986). We thereafter reasoned that if we used random responding as the S1 or “I saw it” task, it would have a similar P300-enhancing effect like a CM task.

The second response (R2) was to be made when S2, the string of numbers appeared on the screen. This response had to be made on a two button response box with the *right* hand. The string of numbers was either “111111,” “222222,” “333333,” “444444”, or “555555.” These were also presented in a random order. A participant had to press the right target button following “111111” and the left non-target button following any other number string. The subjects were also reminded that the presentation of the number was brief and they had to pay attention, otherwise they would miss the number and fail to press the target or non-target buttons correctly.

In the IN group (N = 13, 4 females, Mean age = 19.69, SD = .63) subjects were presented with all irrelevant dates, i.e., there was no birth date presented. In the second part of each trial participants were presented with exactly same stimuli (Targets/Non-targets) as in the SG and CM groups.

In all five CM groups (total N = 60, 12/group, 33 females, Mean age = 20.38, SD = 2.14) the stimuli were exactly the same as in the SG group. However, in the CM groups, subjects were additionally instructed to perform “mental countermeasures” to 1, 2, 3, or 4 (of four) irrelevant stimuli in 1CM, 2CM, 3CM, and 4CM groups respectively. In the 5CM group subjects performed countermeasures to all five stimuli, including the probe. The CM responses were the participant’s silent, mental imaging of his/her first name—CM1, last name—CM2, and names for meaningful people (mother, father, siblings or friends)—CM3, CM4, and CM5. In other words, after a participant saw a to-be-countered irrelevant in the first part of a trial, he/she had to mentally and silently think of one of the names *before* randomly pressing one of the five buttons as R1. Thus we refer to this type of countermeasure as *sequential.* Subjects were instructed to perform countermeasures so that the experimenter could not detect the silent, mental act. In each of the CM groups the to-be-countered irrelevant(s) alternated from subject to subject.

In each of the seven groups, prior to the actual run, participants were given 20–30 practice trials as needed without implementation of any CM(s) to insure they understood and could perform the task. In CM groups, after this practice, the subjects were provided with CM use details and asked to learn the associations between the stimuli and CMs. After the subjects said they were ready, the experimenter would check that the subjects perfectly memorized the associations: The experimenter recited each of the to–be-countered stimuli and required a subject to say out loud the CM associated with this stimulus. This was repeated until there were three consecutive perfect response sets by the subject. The detailed trial structure is presented in Fig. 1 and is identical to the trial structure presented in Rosenfeld and Labkovsky (2010).

## Data Acquisition

EEG was recorded with Ag/AgCl electrodes attached to sites Fz, Cz, and Pz. The scalp electrodes were referenced to linked mastoids. EOG was recorded with Ag/AgCl electrodes above and below the right eye. The diagonal placement of the eye electrodes ensured that both vertical and horizontal eye movements would be picked up, as verified in pilot study and in Rosenfeld et al. (2004, 2008). The artifact rejection criterion was 40 µV. The EEG electrodes were referentially recorded but the EOG electrodes were differentially amplified. The forehead was connected to the chassis of the isolated side of the amplifier system (“ground”). Signals were passed through Grass P511amplifiers with a 30 Hz low pass filter setting, and high pass filters set (3 db) at .3 Hz. Amplifier output was passed to a 16-bit National Instruments A/D converter sampling at 500 Hz. For all analyses and displays, single sweeps and averages were digitally filtered off-line to remove higher frequencies, the digital filter was set up to pass frequencies from 0 to 6 Hz using a “Kaiser” filtering algorithm.

P300 at Pz was measured using the Peak–Peak (p–p) method, which, as repeatedly confirmed in our previous studies, is the most sensitive in P300-based deception investigations (e.g., Soskins et al. 2001): The algorithm searched from 300 to 650 ms for the maximally positive 100 ms segment average. The midpoint of the segment defined P300 latency. Then it searches from this P300 latency to 1,300 ms for the maximum 100 ms negativity. The difference between the maximum positivity and negativity defines the p–p measure. P900 was measured (as is typical in all but deception studies) base to peak (b–p) as the maximum 100 ms segment between 700 and 1,100 ms minus the 100 ms average of pre-S1 EEG.

### Analyses, Error Handling

To determine group effects ANOVAs were run. Any within-subject tests with >1 *df* resulted in our use of the Greenhouse-Geisser (GG) corrected value of probability, p(GG), and the associated epsilon = e value. All artifact trials were discarded so that analyses were done only on artifact free trials.

To ensure that subjects were cooperating with instructions, we monitored S1 random responses during recording. During the recording we verified that subjects were randomizing the choices, and avoiding pressing the same button twice in a row, and not developing any other pattern of button presses. In terms of the S2, the computer monitored accuracy for target and non-target buttons (results below).

### Within Individual Analysis: Bootstrapped Amplitude Difference Method

To determine whether or not the P300 evoked by one stimulus is greater than that evoked by another *within an individual*, the bootstrap method (Efron 1979) was used on the Pz site where P300 is typically largest. This will be illustrated with an example of a probe response being compared with an irrelevant response. The type of question answered by the bootstrap method is: Is the probability more than 90 in 100 that the true difference between the average probe P300 and the average irrelevant P300 is greater than zero? For each subject, however, one has available only one average probe P300 and one average irrelevant P300. Answering the statistical question requires distributions of average P300 waves, and these actual distributions are not available unless one repeats the experiment multiple times which is not feasible. One thus bootstraps these distributions, in the bootstrap variation used here, as follows: A computer program goes through the combined probe-target and probe non-target set (all single sweeps) and draws at random, with replacement, a set of n1 waveforms. It averages these and calculates P300 amplitude from this single average using the maximum segment selection method as described above for the p–p index. Then a set of n2 waveforms is drawn randomly with replacement from the irrelevant set, from which an average P300 amplitude is calculated. The number n1 is the actual number of accepted probe (target and non-target) sweeps for that subject, and n2 is the actual number of accepted irrelevant sweeps for that subject multiplied by a fraction (about .22 on average across subjects in the present report) which reduces the number of irrelevant trials to within one trial of the number of probe trials. The calculated irrelevant mean P300 is then subtracted from the comparable probe value, and one thus obtains a difference value to place in a distribution which will contain 100 values after 100 iterations of the process just described. Multiple iterations will yield differing (variable) means and mean differences due to the sampling-with-replacement process.

In order to state with 90% confidence (the criterion used in all preceding studies, (e.g., Farwell and Donchin 1991; Soskins et al. 2001; Rosenfeld et al. 1991, 2004) that probe and irrelevant evoked ERPs are indeed different, we require that the value of zero difference or less (a negative difference) not be >−1.29 SDs below the mean of the distribution of differences. In other words, the lower boundary of the 90% confidence interval for the difference would be greater than 0. It is further noted that a one-tailed 1.29 criterion yields a $p < .1$ confidence level within the block because the hypothesis that the probe evoked P300 is greater than the irrelevant evoked P300 is rejected either if the two are not found significantly different or if the irrelevant P300 is found larger (*t* tests on single sweeps are too insensitive to use to compare mean probe and irrelevant P300s within individuals; see Rosenfeld et al. 1991).

Finally, in describing the diagnostic accuracy of experiments, we made use of the signal detection theoretical parameter, A′, based on Grier (1971). This is a function of the distance between a ROC curve and the main diagonal of a ROC plot of Hits and False Alarms. It makes no assumptions about the shape or variances of the distributions of the key variables (such as P-I P300 amplitude differences). A′ varies from .5 (null effect) to 1.0 (maximum effect). $A' = \frac{1}{2} + ((y - x)(1 + y - x)/(4y(1 - x)))$; y = hit rate, x = false alarm rate.

## Results: Behavioral

RTs for SG, IN, and CM groups (1–5) are shown in Fig. 2. It appears that SG and IN groups are faster than all CM groups, as expected, and that SG and IN groups are similar except regarding the probe RT, which is slightly elevated in the SG group, as has been previously reported (Farwell and Donchin 1991; Seymour et al. 2000; Rosenfeld et al 2008). It also appears that within the CM groups, countered stimuli have higher RTs than un-countered ones, as in Sokolovsky et al. (2011) and Hu et al. (in press).

Our first major question about RT is: What is the effect of number of irrelevants countered on the RT to countered and non-countered *irrelevants*? The next major question is

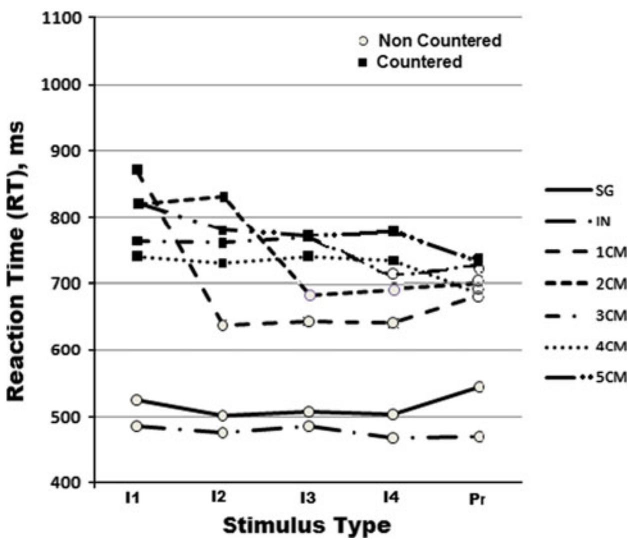


**Fig. 2** RTs as a function of the various stimulus types, groups, and countering status

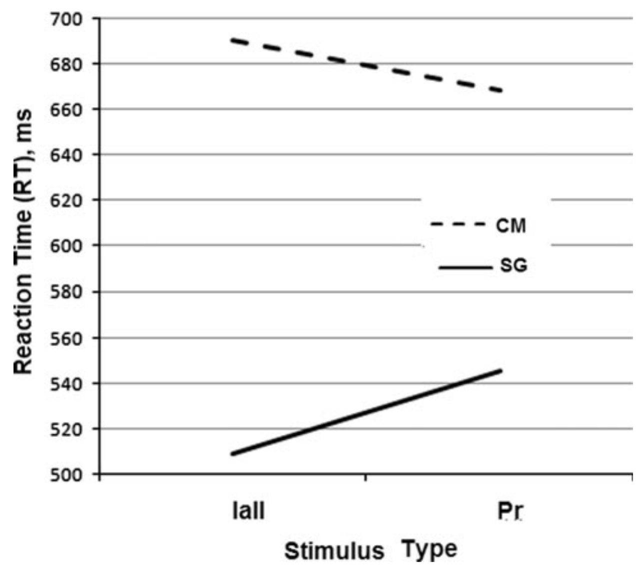


**Fig. 3** RT as a function of stimulus type in which Iall is the combined value for all irrelevant reaction times, countered and not countered, for SG (guilty with no countermeasures) and countermeasure groups (except 5CM) combined

the same one but about *probe* RTs: What is the effect of number of irrelevants countered on the probe RT? These questions first require analysis of Fig. 2, but one section of the figure at a time. Thus we first did a 2-way groups (1-CM, 2-CM, 3-CM, 4-CM, 5-CM) by within-subject stimulus types (I1, I2, I3, I4) ANOVA. For this analysis, it is noted that (a) stimulus types, going from left to right in Fig. 2, involve systematic increases in the ratio of countered to non-countered irrelevant(s) within each vertical column of groups; (b) SG and IN groups are omitted in this analysis as are probe data since there are no CMs used in SG and IN groups, nor against probes in 1-CM through 4-CM groups. The combined irrelevant RTs in these five groups (1CM–5CM) did not differ; $F(4, 54) = .367$, $p > .8$, but the stimulus types did differ, $F(3, 162) = 3.77$, $p(GG) < .03$, and the interaction was also significant, $F(12, 162) = 7.45$, $p(GG) < .001$. The interaction is evident in Fig. 2 as the curves in the CM groups do not resemble one another: In the 1CM and 2CM groups, the countered irrelevant RTs are obviously greater than the RTs to non-countered items, as expected. In the 1CM group, a *t* test on countered versus uncountered RTs was $t(10) = 5.5$, $p < .001$. In the 2CM group, a *t* test on countered versus uncountered RTs was $t(11) = 3.2$, $p < .009$.This trend can be seen in the 3CM group, though it is less obvious (and the *t* test was marginal at $t(11) = 1.9$, $p < .1$). In both the 4CM and 5CM groups, since I1, I2, I3, and I4 are all countered, the RT functions are flat; (so supporting *t* tests were not done as they would only prove a null hypothesis.)

For a more general analysis of these data, we performed a 2-way (group; SG vs. CM) × 2 (stimulus type) ANOVA, where one within-subject variable, stimulus type (probe vs. irrelevant) involved combining all countered and non-countered irrelevant RTs from all four CM groups except the 5-CM group into one average called Iall. We omitted the 5-CM group since the countered probe RT in that group could not have been meaningfully combined with probe data in other groups in which probes were not countered. Figure 3 shows the graph for this analysis. There was a significant effect of group $F(1, 57) = 14.6$, $p < .001$. There was no effect of stimulus type, $F(1, 57) = .01$, $p > .9$, probably related to the significant interaction, $F(1, 57) = 9.8$, $p < .004$, suggested in Fig. 3. To verify that the group effect was not solely carried by the *Iall* RTs (perhaps suggested in Fig. 3), we did a post hoc *t* test comparing just *probe* RTs in CM groups with *probe* RTs in the SG group. T (20.3 pooled variance) = 3.3, $p < .003$. The clear and interesting implication is that the presence of CMs to irrelevants clearly affects probe RT despite the fact that probes are not countered in these four CM groups of Fig. 3. Thus doing CMs causes a general RT slowdown.

To examine the hypothesis of possible systematic changes in RT as a function of the number of countered irrelevants, we performed a 2-way, within-subject stimulus type (probe vs. countered irrelevant vs. non-countered irrelevant) × group (1CM vs. 2CM vs. 3CM) ANOVA; see Fig. 4. (We omitted the 4-CM and 5-CM groups in this analysis as there were no non-countered CMs in those groups. But for the interested reader, in the 4CM group, the probe RT = 685.4, and the countered average for the 4 countered irrelevants was 736.3. In the 5CM group, the probe RT = 731.0, and the countered average for the 4 countered irrelevants was 787.6.) Again there was no main

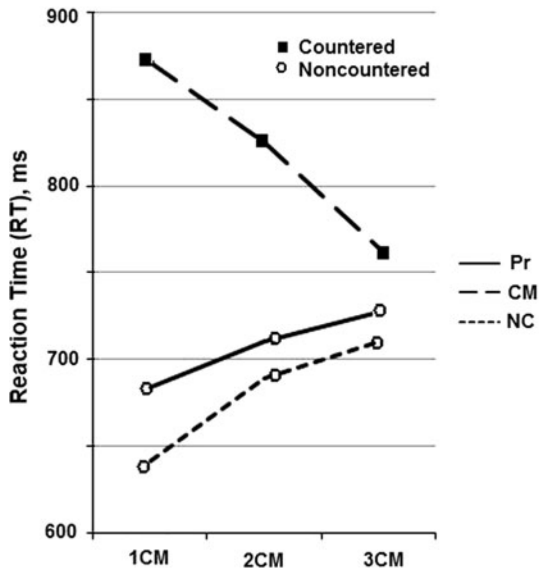


**Fig. 4** RT as a function of 3 countermeasure groups (1CM, 2CM, 3CM) and stimulus types; probes (Pr), countered irrelevants (CM) and not countered irrelevants (NC)

effect of group; F (2, 32) = .01, $p > .9$, probably due to the interaction: F (4, 64) = 3.9, p(GG) < .02), evident in Fig. 4. There was also a large main effect of stimulus type, F (2, 64) = 28.7, p(GG) < .001. Figure 4 shows probe and non-countered RTs increasing with numbers of countered irrelevants, as expected, but the less intuitively obvious, decreasing function for countered irrelevants. Post-hoc tests within each stimulus type failed to reach significance. The inverse relation between RT and number of countered irrelevants may be related to preparation effects, meaning that one is more likely to be expecting and ready to execute a CM as the number of CMs increase.

## Results: ERPs

Figure 5 shows grand average ERPs (probe and Iall) at all sites, in all guilty groups (SG and all CM) used in the study. The IN results are not shown here as there there were no P300s or P900s in this group (as seen in Figs. 6, 7). The probe P300s in the Pz traces are indicated by thick vertical lines below zero uV axes at P300 latency (500–600 ms) to these alphanumeric stimuli. P900s in Fz and Cz traces, when evident, are indicated with thick vertical lines above the zero uV axes at 850–950 ms. Prominent P300s are seen in the Pz traces of the SG, 1CM, 4CM, and 5CM groups, with smaller P300s in the 2CM and 3CM groups. On the other hand, prominent P900s are seen in Fz and Cz traces of 2CM and 3CM groups, with no suggestion of P900 in SG and 5CM groups.

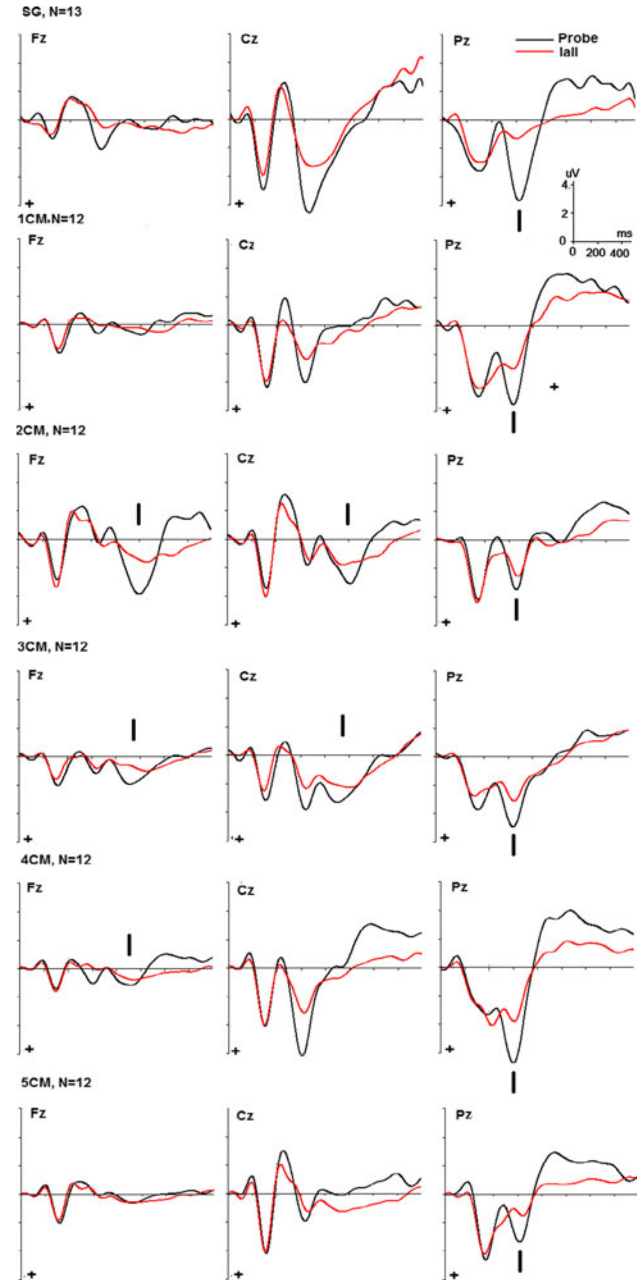


**Fig. 5** Grand averaged waveforms for all guilty groups, as a function of site and stimulus type with all irrelevants combined. *Positive down* as shown. *Vertical lines* below *x*-axes in Pz waveforms show P300. *Vertical lines* above *x*-axes in in some Fz and Cz waveforms show P900

## P300 Group Data Analysis

Figure 6 shows the computer calculated peak-peak P300 Pz amplitudes for all seven groups studied here. There appear to be large probe-irrelevant differences in all groups but the IN group, as expected. The largest differences appear to occur in the SG and 4CM groups. The large difference for the SG group is expected since there are no CMs, so the

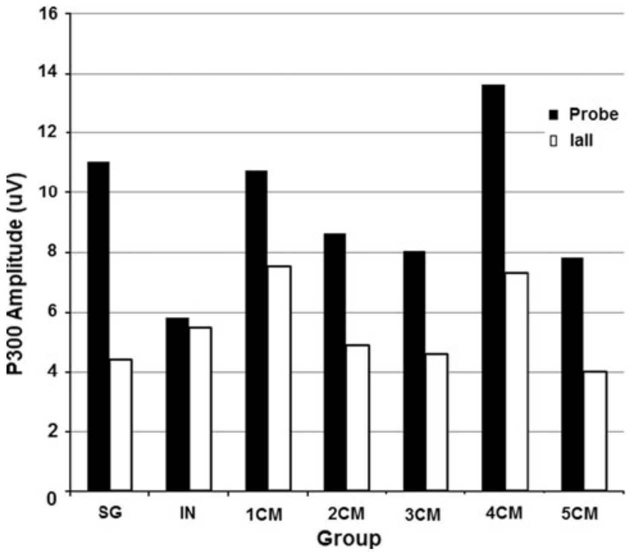


**Fig. 6** Computer calculated p–p P300 probe and Iall amplitudes at Pz for each of the 7 groups

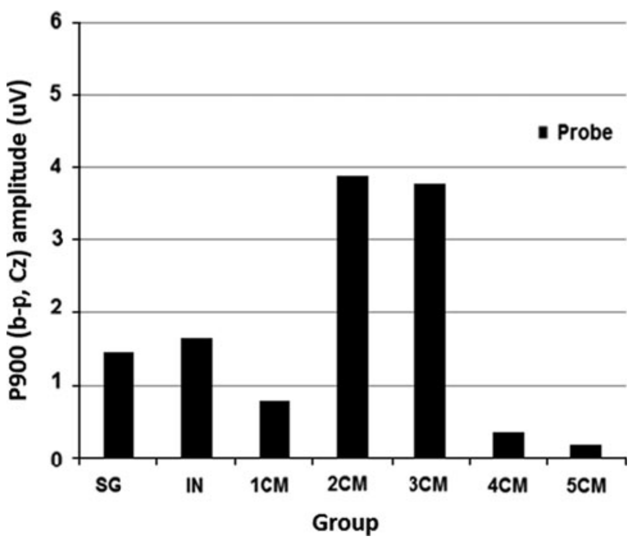


**Fig. 7** Computer calculated b–p P900 probe amplitudes at Cz for each of the 7 groups

irrelevant P300 is small but the probe P300 is large. The large difference for the 4CM group is expected on the basis of the omit effect (Meixner and Rosenfeld 2010) described above. The first confirmatory analysis run on these data was a 2 way (stimulus type: probe vs. Iall, within subject) × (groups: all 7 groups shown in Fig. 6, between subjects) ANOVA. The effect of groups was significant, F (6, 79) = 3, 22, $p < .008$); as were the effects of stimulus type, F (1, 79) = 253.3, $p < .001$; and interaction, F (6, 79) = 10.21, $p < .001$. The groups effect reflected the P300 (combined probe and Iall) differences across groups and the interaction reflected the *differences* among these probe-irrelevant differences across groups described above.

Post-hoc ANOVAS followed up the effects described above. First, to confirm similarities of probe versus Iall amplitudes within the 4CM and SG groups, and secondly, within the 1CM, 2CM, 3CM, and 5CM groups, a first, 2 way (stimulus type: probe vs. Iall) × (groups: the four just cited) ANOVA found no effect of groups, F (3, 44) = 1.68, $p > .19$, nor of interaction, F (3, 44) = .097, $p > .9$. The difference between probe and Iall remained significant, F (1, 44) = 165.96, $p < .001$. Second, we show that the SG and 4CM groups have similarly large probe—Iall differences in a 2 way (stimulus type: probe vs. Iall) × (groups: SG vs. 4CM) ANOVA. The groups effect, favoring the 4CM group, was not significant, F (1, 23) = 2.99, $p = .096$ and the interaction was not significant, F (1, 23) = .084, $p > .77$. The probe versus Iall effect remained significant, F (1, 23) = 139.1, $p < .001$. Finally, to support the visual observation of the SG and 4CM groups having the largest probe versus Iall difference, a post-hoc *t* test compared the mean differences in these two combined groups versus the combined mean differences in the 1CM, 2CM, 3CM, and 5CM groups, yielding a significant t(37.2 separate variance *df*) = 4.61, $p < .001$, showing a larger P300 difference in the SG and 4CM groups.

## P900 Group Data Analysis

Figure 7 shows computer calculated probe P900s (b–p) recorded from Cz in bar graph format for all seven groups studied here. The comparable Fz data (not shown here) look similar. It is recalled that our hypothesis about P900 in the present studies is that it should be seen in the probe ERP (and occasionally in the non-countered irrelevant P900), but not in the countered irelevant P900. Consistent with Figs. 5, 7 shows that there appear to be large P900 responses in the probe waveforms in the 2CM and 3CM groups, but not in other groups, as hypothesized. A first approach to analysis of these Cz data involved a 2 way (one stimulus type; probe) by 7 (groups) ANOVA. The effect of groups was significant (F [6, 79] = 3.84, $p < .003$). To confirm the fact that P900 is larger in the 2CM and 3CM groups, we did a 1 (stimulus type; probe) by 2 (group; combined 2CM and 3CM vs. other 4 guilty and CM combined groups) ANOVA, yielding the expected main effects of groups; F (1, 71) = 21.2, $p < .001$.

We did the same analyses at Fz: In the 1 (stimulus type) × 7 (groups) ANOVA, the group effect was significant: F (6, 79) = 9.4, $p < .001$. Doing the ANOVA on combined 2CM and 3CM versus 1CM, 4CM, and 5CM groups, (as with Cz data), the group effect was F (1, 71) = 29.7, $p < .001$.

### Individual, P300-Based Detection Rates

The individual detection rates in all groups are shown in Table 1. These rates are based on bootstrapping the probe minus irrelevant Pz P300 (p–p) amplitude difference at a confidence interval = 0.9. It is clear that detection rates were always >90% with <8% false positives, yielding Grier (1971) A′ values all >.95.

### Detection of Individual Countermeasure Use with RT

Since we saw here as we did previously (Rosenfeld et al. 2004, 2008, and Rosenfeld and Labkovsky 2010) that sequential countermeasure use elevates irrelevant RT as a strong group effect, we wondered to what extent elevated irrelevant RT might enable us to detect *individuals* attempting to make covert responses to irrelevants. The main concern with a countermeasure user is that his average Iall P300 response will be indistingushable from his probe P300 response as one expects in an innocent subject. An irrelevant (Iall) RT-based classification scheme for countermeasure use should thus be based on known innocent subjects also known not to be using countermeasures. Our innocent subjects have a mean Iall of 479 ms ± 141 ms. Table 2. shows the numbers of diagnosed countermeasure users (with associated Grier 1971 A′ values) in the various groups based on the Iall mean + three SD functions: 1 SD, 1.3 SD, and 2 SD, which should respectively include about 84, 90, and 98% of true innocent values. The 1.3 SD value falsely labels 1 subject each in the IN and SG groups, while yielding respectable countermeasure use detection rates in the countermeasure groups—whose members are mostly detected as guilty based on P300 anyway.

**Table 2** RT-based countermeasure-user diagnoses with Grier (1971) A′ values, based on 3 criteria, all groups

| Group (n) | Criterion = Iall mean (479 ms)+ | | |
|---|---|---|---|
| | 1 SD = 620 ms | 1.3 SD = 662 ms | 2 SD = 761 ms |
| 1CM (12) | 8 (A′ = .87) | 8 (A′ = .87) | 5 (A′ = .79) |
| 2CM (12) | 9 (A′ = .90) | 7 (A′ = .85) | 5 (A′ = .79) |
| 3CM (12) | 9 (A′ = .90) | 9 (A′ = .90) | 5 (A′ = .79) |
| 4CM (12) | 8 (A′ = .87) | 8 (A′ = .87) | 7 (A′ = .85) |
| 5CM (12) | 10 (A′ = .93) | 9 (A′ = .90) | 6 (A′ = .82) |
| SG (12) | 3 (A′ = .90) | 1 (A′ = .95) | 0 (A′ = .98) |
| IN (11) | 1 (A′ = .95) | 1 (A′ = .95) | 1 (A′ = .95) |

### Discussion

The present results extend earlier reports (Rosenfeld et al. 2008; Rosenfeld and Labkovsky 2010; Winograd and Rosenfeld 2011) on the P300-based, complex trial protocol for detecting concealed information, by showing that the version of the protocol using one probe and four irrelevant stimuli, remains resistant to any possible fraction of countered irrelevants, (Table 1) as long as the *sequential* countermeasure is executed prior to the stimulus acknowledgement. The present results indicate, however, that the large probe-irrelevant P300 difference that one sees (Fig. 5) in SG groups or countermeasure groups with proportions of countered irrelevant stimuli <.5 or >.75, does *not* occur in countermeasure groups with proportions of countered irrelevant stimuli varying from .5 to .75 (Fig. 5). With such middle proportions (2/4, 3/4) of countered irrelevants, although the accuracy rates in the present study were 100%, the p–p probe amplitudes were less large than those in the 1CM and 4CM groups.

Much of this effect, as can be seen in Fig. 5, is due to the fact that the negative overshoot component (utilized in the p–p measure of P300) that represents the recovery of P300 from its positive peak is pulled down by the occurrence of P900—which is also prominent in the 2CM and 3CM groups (mostly at Fz and Cz, but somewhat reflected at Pz where P300 is measured).

Average irrelevant RT continues [as in Rosenfeld et al. (2008), Rosenfeld and Labkovsky (2010), and Winograd and Rosenfeld (2011)] to be a good indicator of *sequential* countermeasure use, as Fig. 3 shows an average difference of 180 ms between irrelevant RT in SG versus IN countermeasure-using subjects. Actually, even if one does not know which stimulus is the probe, as in some field situations (Meixner and Rosenfeld 2011), Fig. 3 shows that the average RT for all five stimuli is about 160 ms greater in a CM group than for an SG group.

On the other hand, Sokolovsky et al. (2011) showed that *simultaneous* countermeasures, executed *at the same time as* R1, the "I saw it" stimulus acknowledgement, renders RT ineffective as a countermeasure use indicator in a situation involving one probe and four irrelevants, such as the present protocol. P300-based detection of concealed information still remained good (85%) in that situation. Also fortunately, Hu et al. (in press) have shown that increasing the number of irrelevants to eight restores the ability of RT to index even *simultaneous* countermeasure use while yielding a P300-based Grier (1971) A′ of .93 overall.

Finally, P900, may prove a helpful additional index of counterneasure use. It continues to appear in response to uncountered probes at Fz and Cz in situations involving middle fractions of countered irrelevant stimuli, as previously reported by Rosenfeld and Labkovsky (2010). We continue to hypothesize that its appearance following a probe represents the probe's definitive signal to the subject that no further stimuli in the trial will require a (countermeasure) response. We do not see P900 in 1CM, 4CM, and 5CM groups perhaps because in these less ambiguous situations, the P300 component is more prominent due to less task demand, as noted above, and its enhanced prominence obscures P900. Clearly, more research in delineating the functional significance of P900 is in order. However, it is fortunate for would-be detectors of concealed information that when P900 is not evident, it is least needed as a supplementary index of countermeasure use, since the probe-irrelevant P300 differences are largest when P900 is not easily seen.

## References


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

Ben-Shakhar, G., & Elaad, E. (2002). The Guilty Knowledge Test (GKT) as an application of psychophysiology: Future prospects and obstacles. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 87–102). London: Academic Press.

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.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26.

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

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

Grier, J. B. (1971). Non-parametric indexes for sensitivity and bias: Computing formulas. *Psychology Bulletin, 75*, 424–429.

Honts, C., Amato, S., & Gordon, A. (2001). Effects of spontaneous countermeasures used against the comparison question test. *Polygraph, 30*(1), 1–9.

Honts, C., Devitt, M., Winbush, M., & Kircher, J. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology, 33*, 84–92.

Hu, X., Hegeman, D., Landry, E., & Rosenfeld, J. P. (in press). Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection. *Psychophysiology*.
.

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

Meixner, J., Haynes, A., Winograd, M., & Rosenfeld, J. P. (2009). Assigned versus random, countermeasure-like responses in the P300 based Complex Trial Protocol for Detection of Deception: Task demand effects. *Applied Psychophysiology & Biofeedback, 34*, 209–221.

Meixner, J. B., & Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300-based concealed information test. *Psychophysiology, 47*, 57–65.

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.

National Research Council. (2003). *The Polygraph and Lie Detection*. Committee to review the scientific evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991). An ERPbased, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology, 38*, 319–335.

Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, eventrelated potential-based guilty knowledge test. *International Journal of Neuroscience, 24*, 157–161.

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*(6), 1002–1010.

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

Seymour, T. L., Seifert, C. M., Mosmann, A. M., & Shafto, M. G. (2000). Using response time measures to assess ''guilty knowledge. *Journal of Applied Psychology, 85*, 30–37.

Sokolovsky, A. W., Rothenberg, J., Labkovsky, E., Meixner, J. B., & Rosenfeld, J. P. (2011). A novel countermeasure against the reaction time index of countermeasure use in the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology, 81*, 60–63.

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

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