Psychophysiology, 49 (2012), 85–95. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2011.01286.x

Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection

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

Abstract

We previously found that simultaneously executing a mental countermeasure and an explicit required response impairs reaction time (RT)-based detection of countermeasure use in a P300- based concealed information test. To address this issue, we increased the numbers of irrelevant stimuli to eight, and manipulated the proportions of to-be-countered irrelevant stimuli from 25% to 50% to 75% in three groups. Results: Based on P300 data, 100% of the simple guilty (no countermeasure use) and 92% of the innocent subjects were correctly identified as having or not having concealed information. In the countermeasure groups, detection rates varied from 71% to 92% across the different groups. Notably, in the present study with eight irrelevant items, simultaneous countermeasure use was indicated by elevated RT in the 50% and 75% countermeasure proportion groups, which it was not, previously, with 50% (two) countermeasures and four irrelevants.

Descriptors: P300, Deception detection, Concealed information test, Countermeasures, Complex trial protocol, Credibility assessment

With renewed intensity since September 11, 2001, there have been enormous efforts expended by governments and universities to continue the century-old development of an accurate deception test based on sound scientific principles. Both polygraph protocols using measurements of autonomic nervous system activity (the Comparison Question Test (CQT) and the Concealed Information Test (CIT)) have been alternatively advocated and criticized, as summarized in the recent report by the National Research Council of the National Academy of Sciences (National Research Council, 2003). Among the problems with polygraphy raised by the National Research Council report is its potential susceptibility to countermeasures. "Countermeasures are anything that an individual might do in an effort to defeat or distort a polygraph test" (Honts, Devitt, Winbush, & Kircher, 1996, p. 84). The National Research Council report stated: "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). More specifically, countermeasures are effective against both the polygraphic CQT (Honts & Amato, 2002; Honts, Amato, & Gordon, 2001), as well as against the polygraphic CIT (Ben-Shakhar & Dolev, 1996; Honts et al., 1996).

The authors thank Gershon Ben-Shakhar for his many helpful suggestions for improving the earlier versions of this article.
Address correspondence to: Xiaoqing Hu, Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208. E-mail: xiaoqinghu@u.northwestern.edu

It was anticipated that when the relatively fast P300 event-related potential (ERP) response to test items was introduced as a new index of their recognition in a CIT (Allen, Iacono, & Danielson, 1992; Farwell & Donchin, 1991; Rosenfeld, Angell, Johnson, & Qian, 1991; Rosenfeld et al., 1988), the countermeasure issue would be resolved. Thus, in agreement with Ben-Shakhar and Elaad, (2002), the developer of the CIT, Lykken (1998, p. 293), expressed the hope that "because such potentials are derived from brain signals that occur only a few hundred milliseconds after the GKT (Guilty Knowledge Test) alternatives are presented . . . it is unlikely that countermeasures could be used successfully to defeat a GKT derived from the recording of cerebral signals." Unfortunately, however, Rosenfeld, Soskins, Bosh, and Ryan, (2004) and Mertens and Allen (2008) showed that the original form of the P300-based CIT was vulnerable to countermeasures, prompting development of a novel P300-based protocol that has thus far resisted previously effective countermeasures (Mertens & Allen, 2008) in three new studies (Rosenfeld & Labkovsky, 2010; Rosenfeld et al., 2008; Winograd & Rosenfeld, 2011).

Indeed, this novel complex trial protocol has so far been the only physiologically based direct or indirect deception testing protocol reported that is resistant to countermeasures and, additionally, provides a simple reaction time (RT) index of the use of a countermeasure by a subject. Thus, this protocol not only identifies guilty knowledge recognition in the face of verbal denial, but additionally identifies attempts to beat the test. This likely constitutes useful independent evidence of a subject's criminal complicity. Especially in the rare case in which a subject succeeds in beating the test by not showing the enhanced P300 indicator of guilty knowledge recognition, the subject's RT index may still give away his or her attempt at noncooperation—useful information for enforcement officials since innocent subjects have no incentive to counter the test. Therefore, it was most disappointing to identify a novel countermeasure strategy (Sokolovsky, Rothenberg, Meixner, & Rosenfeld, 2011) that defeated our previously effective RT index of countermeasure use. This countermeasure strategy was anticipated neither by ourselves nor by several previous reviewers.

In the original report of the complex trial protocol, for guilty participants not using countermeasures, as well as for innocent participants, the first stimulus acknowledgement (or "I saw it") response, which involved pressing a button upon seeing either a probe (i.e., a rare concealed information item) or irrelevant stimulus (i.e., frequent filler items), was relatively fast (short RT), for it did not involve any countermeasure selection process. Countermeasure users, however, were instructed to do the countermeasure before pressing the "I saw it" button. This sequential execution of countermeasure first and then the "I saw it" response made the countermeasure users' irrelevant RT become significantly longer due to the countermeasure selection process (e.g., the participants must detect that the specific irrelevant requires a countermeasure and then choose the corresponding countermeasure). Elevation of RT for the "I saw it" response was thus an effective countermeasure use indicator (Rosenfeld et al., 2008; Winograd & Rosenfeld, 2011). The recently demonstrated simultaneous countermeasures strategy just noted (Sokolovsky et al., 2011), however, involves a participant's execution of a countermeasure at exactly the same time as she presses the "I saw it" button. Thus, the countermeasure-use indicator based on the "I saw it" RT became ineffective (Sokolovsky et al., 2011).

To address this issue, we increased the number of irrelevant stimuli from four in Sokolovsky et al. (2011) to eight. The logic was that increasing the number of irrelevants would increase the task demands of executing countermeasures; that is, countermeasure users would now need to hold more irrelevant stimuli in working memory and to classify more irrelevant stimuli as either countermeasure-assigned or not. This increased task demand is hypothesized to increase the RT of the "I saw it" response, making the simultaneous countermeasures' users detectable and thus restoring our countermeasure-use indicator.

In addition, we systematically manipulated the number of irrelevant stimuli that the guilty participants had to counter so as to examine the complex trial protocol's resistance to various countermeasure proportions. In the earlier complex trial protocol studies, subjects performed countermeasures to all the irrelevant stimuli but not to the probe, and this made the probe the only stimulus that did not require a countermeasure response. This unique probe assignment (i.e., withholding the countermeasure response) probably increased the salience of this probe stimulus and its corresponding P300 amplitude (this was shown in a study in which one of five irrelevants was the only stimulus not requiring the countermeasure, and it alone elicited a large P300; Meixner & Rosenfeld, 2010). This finding implied that the more effective countermeasure to defeat the complex trial protocol would involve doing countermeasures to a smaller fraction of irrelevant stimuli. To explore this question, we manipulated the numbers of countermeasures used from 25% to 50% and 75% of all irrelevant stimuli, which allowed us to observe for the first time the P300's responsiveness over a representative range of simultaneous countermeasures. It also allowed us to observe how these different countermeasure proportions affect the RT index of simultaneous countermeasure use.

Methods

Participants

Seventy-two participants at Northwestern University were recruited from the Introductory Psychology subject pool and were given course credit for participation. Nine participants were excluded due to defective Pz electrodes or excessive eye movement artifacts. Thus, data from 63 participants were used for the final analysis (age range 18–22, 30 males). Participants provided informed consent prior to the experiment. All participants were right-handed and had normal or corrected vision. The study was approved by the Northwestern Institutional Review Board.

Procedure. Participants were randomly assigned to one of five groups/conditions: we had 12 participants in the simple guilty group (these are guilty subjects not using countermeasures), 12 participants in the innocent control group, 13 participants in the 2-countermeasure group, 12 participants in the 4-countermeasure group, and 14 participants in the 6-countermeasure group. Before the experiment began, all participants except those in the innocent group were asked to provide their hometown name, to be used as a probe in the experimental run. Next, they were provided with a list of possibly irrelevant city or town names to see whether any of these were personally meaningful. Only those stimuli with no personal significance were selected as irrelevant. All stimuli appeared in 1 cm tall white font on a black background. The stimulus display was about 1 m from the participants' eyes.

The trial structure was similar to that of Rosenfeld et al. (2008). Each trial began with a 100-ms baseline period for the recording of prestimulus electroencephalogram (EEG). The probe or irrelevant was then presented on the center of the screen for 300 ms. Following a randomly varying interstimulus interval lasting 1400–1700 ms, the target/nontarget stimulus was presented also for 300 ms. The target task was used solely to force attention so no data from it were collected. There were 360 trials in total, consisting of a probe and eight irrelevants, repeated 40 times each, for a total of 40 probes and 320 irrelevants. The experiment lasted for approximately 30 min.

In the simple guilty group, participants first saw either a probe or one of the eight irrelevants, for a probe probability of 1/9. Participants were told to respond randomly on a five-button box by pressing one of the five buttons chosen randomly with their left hand as soon as they saw the stimulus. This was the stimulus acknowledgement or the "I saw it" response. They were warned that the experimenter would pause the experiment about every 20–40 trials and ask them to repeat aloud the probe or irrelevant stimulus just presented. Failure to correctly identify more than one stimulus was indicative of inattentive noncooperation, and the participant data would be dropped. However, there was no attrition for this reason. The probe/irrelevant was followed about a second later by a target or nontarget number string (11111, 22222, 33333, 44444, and 55555). Participants were asked to make a target/nontarget decision with their right hand upon seeing the string of numbers. If the string of numbers was "11111," they were told to press the right button (target) of the response box with their middle finger, and to press the left button of the response box with their index finger if the stimulus was any other string of numbers (nontarget). The target and nontarget occurred at an equal probability following probe or following each irrelevant.

In the innocent group, participants were presented with all irrelevant stimuli; that is, their hometown was not one of the stimuli seen. All other instructions were the same as in the simple guilty group.

In the 2-countermeasure group, the instruction and task were the same as for the simple guilty group except that the participants were told to additionally execute two specific mental countermeasures to a designated two of the eight irrelevants. In particular, participants were instructed to mentally say their first name to one of the to-be-countered irrelevants and their last name to the other to-be-countered irrelevant. Participants were instructed to mentally say the names upon encountering the corresponding to-be-countered irrelevant at the same time as they pressed one random response button. This is the simultaneous countermeasure.

In the 4-countermeasure group, all procedures were the same as in the 2-countermeasure group except that the number of countermeasures was increased from two to four. Along with the participant's first and last name, the participant's father's first name and mother's first name were both assigned as countermeasures to separate additional irrelevant stimuli.

In the 6-countermeasure group, the countermeasure number was increased to six. Participants were told to do the four countermeasures named above as well as two additional countermeasures: the participant's middle name and mother's maiden name were assigned as countermeasures to two separate additional irrelevant stimuli.

All to-be-countered irrelevant stimuli were assigned one and only one countermeasure (one and only one name) to be performed every time the particular irrelevant appeared. All groups other than the innocents were instructed to conceal hometown names, so all were guilty of this intentional concealment.

Data Acquisition

EEG was recorded using Ag/AgCl electrodes attached to three midline sites: Fz, Cz, and Pz. Scalp electrodes were referenced to linked mastoids. Electrode impedance was kept below 5 kΩ. Electrooculogram (EOG) was recorded differentially via Ag/AgCl electrodes placed diagonally above and below the right eye to record vertical and horizontal eye movements as well as eye blinks. EOG voltages were called artifacts if above 50 μV, and all data from associated trials were rejected. The forehead was connected to the chassis of the isolated side of the amplifier system ("ground"). Signals were passed through Grass P511K amplifiers with a 30-Hz low-pass filter and 0.3-Hz high-pass filter (3 db). Amplifier output was passed through a 16-bit A/D converter sampling at 500 Hz. After initial recording, single sweeps and averages were digitally filtered off-line to remove higher frequencies; the digital filter was set up to pass frequencies from 0 to 6 Hz (3-db point).

Analysis Methods

P300 amplitude at Pz was measured using the peak-peak (p-p) method, which is found to yield more accurate diagnosis than the traditional baseline-peak method in P300 Concealed Information Tests (Soskins, Rosenfeld, & Niendam, 2001): the algorithm searches from 400 to 700 ms for a maximal positive 100-ms segment average. The midpoint of the segment is defined as the P300 latency. Next, the algorithm searches from this P300 latency to 1300 ms for the maximum average 100-ms negativity. The difference between the maximal positive segment and the maximal negative segment is defined as the P300 p-p amplitude.

All artifact-contaminated trials were rejected, so both ERP and RT analyses were conducted on only artifact-free trials.

Within-Individuals Bootstrap Analysis

In addition to the group effects, classification of each participant as guilty or innocent was also a major concern. In particular, we were interested in the detection rate among the countermeasures users. To determine whether a given participant was concealing information or not, the key dependent measure is the difference of P300 amplitude between probe and average of all irrelevants (Iall). Since there is no actual average P300 distribution available, we adopted the repeated random sampling bootstrap method to draw artifact-free samples with replacement from the probe or irrelevant category (Wasserman & Bockenholt, 1989). With iteration, this method allows us to obtain multiple bootstrapped averages to generate a bootstrapped simulated distribution of average P300 waves. Here, the p-p P300 amplitude at the Pz site was used since Pz is where the P300 is typically the largest. The procedure worked as follows: First, a computer program draws randomly, with replacement, from all accepted probe single sweeps, a set of sweeps of the same size as the original probe sweep set and averages them so as to obtain one individual probe average. Second, the same program draws randomly, with replacement, a number of accepted irrelevant single sweeps (both countered and noncountered), which is equal to the number of probes so as to obtain one bootstrapped average of irrelevants. Third, the average irrelevant P300 is subtracted from the average probe P300 to obtain a difference score. These steps are iterated 100 times to obtain 100 such difference scores.

To determine whether the probe P300 is larger than the irrelevant Iall P300 with 90% confidence, at least 90 out of 100 difference scores should be above zero. Thus, any participant whose probe is found to be larger than his irrelevant in more than 90 iterations is classified as guilty.

Finally, to determine the overall discriminative ability of the complex trial protocol, we take advantage of the signal detection theoretical parameter A' based on Grier (1971). Here, $A' = 0.5 + \{(y - x) \times (1 + y - x)/[4 \times y \times (1 - x)]\}$, in which the y means hit rate and x means false positive rate. The value of A' varies from .5 (representing no detection efficiency at all) to 1 (indicating perfect discrimination between guilty and innocent).

Results

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

Manipulation Check

To ensure that the participants followed the instructions and that our manipulations were successful, we examined the RT and P300 in the three countermeasure groups. For this analysis, the irrelevant RTs and P300s were subdivided into irrelevant-countered and irrelevant-noncountered RTs and P300s (see Figures 1, 2). From Figure 1, countermeasure use could be visually identified via the elevated irrelevant-countered RT across the three groups. These observations were statistically tested with a mixed, 3 × 3 ANOVA with the three groups as the levels of the between-subject variables and stimulus type as the within-subject variable (3 stimulus types). Results showed that the effect of stimulus type on RT was significant: $F(2,72) = 9.06$, $p(\mathrm{GG}) < .01$, $\eta^2 = .201$. Planned comparisons showed that the irrelevant-countered had a significantly longer RT than both the probe ($t(38) = 3.066$, $p < .01$) and irrelevant-

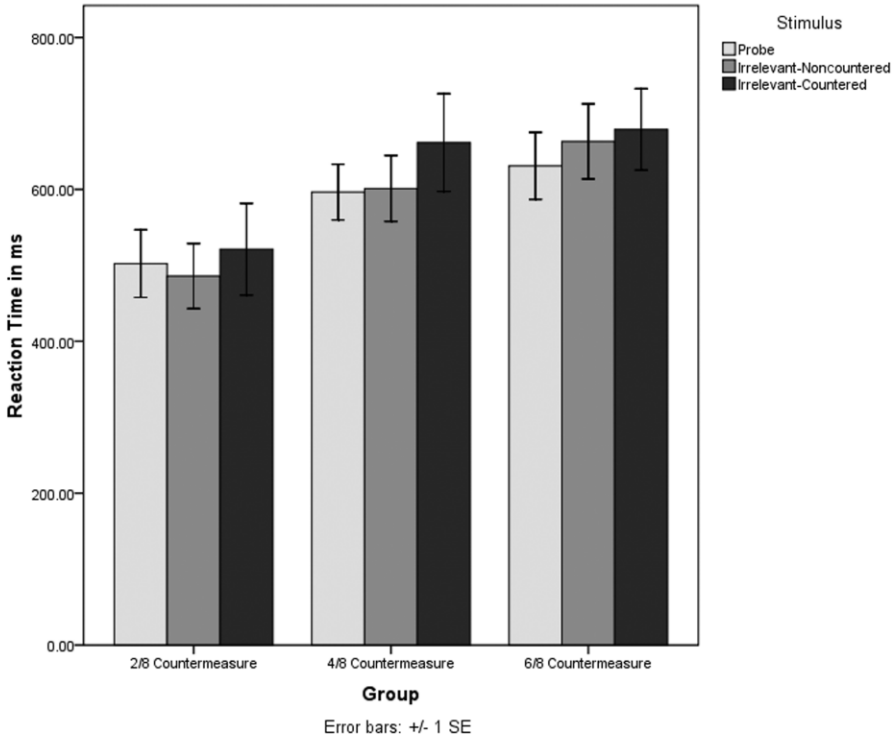

Figure 1. Reaction time (in milliseconds) of the first "I saw it" response for probe, irrelevant-countered, and irrelevant-noncountered in the 2-, 4-, and 6-countermeasure groups. The error bar stands for one standard error (*SE*).

noncountered ($t(38) = 3.135$, $p < .01$), whereas there was no difference between probe and irrelevant-noncountered ($t(38) = 1.126$, $p > .2$). No other main effect or interaction was significant (all $p > .05$). These data help support the fact that subjects were countering the appropriate irrelevants, leading to longer RTs with increasing numbers of countered irrelevants.

Next, we analyzed the Pz-P300 across the three countermeasure groups (see Figure 2a, b). From Figure 2b, it is observed that the p-p P300 to the probe was the largest among three stimulus types, even with as many as 6 of 8 countered irrelevants. To confirm these observations, we ran a mixed 3 × 3 ANOVA using the three countermeasure groups as the between-subject variable and the three stimulus types as the within-subject variable. Results showed a strong main effect of stimulus type: $F(2,72) = 70.248$, $p < .001$, $\eta^2 = .661$. Planned contrasts showed that the probe elicited a larger P300 than the irrelevant-countered ($t(38) = 6.359$, $p < .001$) and the irrelevant-noncountered ($t(38) = 12.065$, $p < .001$). Moreover, the irrelevant-countered elicited a larger P300 than the irrelevant-noncountered ($t(38) = 3.549$, $p < .001$), suggesting the participants' mental countermeasures did elicit P300, which also supports the fact that countermeasures were being executed to the correct irrelevants. Group also showed a significant main effect: $F(2,36) = 4.772$, $p < .05$, $\eta^2 = .21$. Planned comparison between groups showed that the P300 in the 2-countermeasure group was larger than that in the 6-countermeasure group ($t(25) = 2.535$, $p < .05$); no other contrast difference was found. The interaction between stimulus type and group was significant: $F(4,72) = 5.809$, $p(GG) < .001$, $\eta^2 = .244$.

To further analyze this interaction, we compared the probe, irrelevant-countered, and irrelevant-noncountered P300s in the three countermeasure groups separately. In the 2-countermeasure group, the probe and irrelevant-countered elicited a larger P300 than irrelevant-noncountered ($p < .05$), confirming that, in this 2-countermeasure group, the rarer, meaningful stimuli were eliciting P300. In the 4-countermeasure group, although probe P300 was still larger than the irrelevant-noncountered P300 ($p < .05$), the irrelevant-countered did not differ from irrelevant-noncountered ($p > .7$); in the 6-countermeasure group, the P300 for probe was larger than both irrelevant-countered and irrelevant-noncountered (both $p < .01$), whereas there was no difference between irrelevant-countered and irrelevant-noncountered ($p > .9$).

The above analyses showed that participants in countermeasure groups followed the instructions, as indicated by both the RT and the P300. It also appears that the difference between countered and noncountered irrelevant P300 declines systematically with number of countered irrelevants, suggesting that increasing the number of countermeasures increases workload resulting in reduced P300 amplitude, as showed by Kramer, Sirevaag, and Brauner (1987). It was also to be expected that the countered irrelevant P300 declines as the number of countermeasures increases since the probability of the countered irrelevant increases with number of countered irrelevants, and in view of the inverse relationship of stimulus probability and P300 amplitude.

Main Hypothesis Testing

We hypothesized that the RT of the countermeasure users should be larger than the RT of the simple guilty and innocent participants. The RT and P300 (at Pz) of probe and Iall (average of all irrelevants) from simple guilty, three countermeasure groups and innocent group are presented in Figures 3 and 4. From Figure 3, it is observed that the probe had a slower RT than Iall in the simple

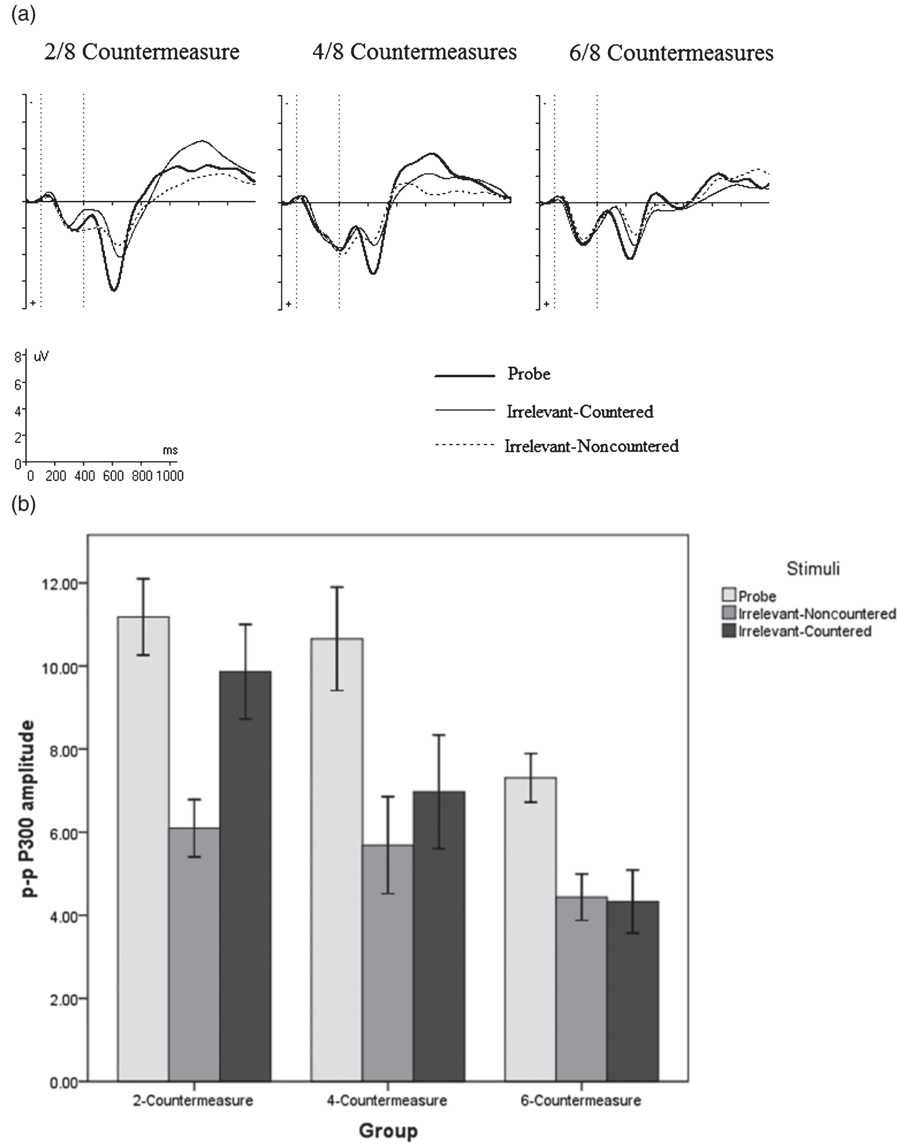

Figure 2. (A) Grand average peak-peak P300 at Pz for probe, irrelevant-countered, and irrelevant-noncountered in the 2-, 4-, and 6-countermeasure groups. The first tick marker indicates the onset of the stimulus presentation; the second tick marker indicates the 300 ms after the stimulus presentation. (B) Bar graph of the same data. The error bar stands for one standard error (*SE*).

guilty group, as seen previously (e.g., Seymour & Kerlin, 2008; Seymour, Seifert, Mosmann, & Shafto, 2000). However, this difference was reversed in the 4- and 6-countermeasure groups: The Iall RT was slower than the probe RT here, probably due to the task demand of executing countermeasures (Rosenfeld & Labkovsky, 2010; Rosenfeld et al., 2008). There was no difference in the less demanding 2-countermeasure group. In the innocent group, there was no noticeable difference between probe and Iall as expected since in this condition, the probe was just another irrelevant. These observations were tested statistically using a 5×2 mixed ANOVA with groups (5 levels: simple guilty vs. 2- vs. 4- vs. 6-countermeasure vs. innocent group) as a between-subject variable and stimulus type (2 levels: probe vs. Iall) as a within-subject variable. Results showed that the effect of stimulus type was not significant ($F(1,58) = 1.341$, $p > .25$, $\eta^2 = .023$), probably related to the group × stimulus type interaction described below. The group effect was significant: $F(4,58) = 3.477$, $p < .02$, $\eta^2 = .193$. *Post hoc* Tukey tests revealed that the 6-countermeasure group was associated with a significantly longer RT than in the simple guilty group ($p < .05$). No other two groups' difference was significant ($p > .05$).

Another important finding was the significant interaction between group and stimulus type: $F(4,58) = 5.371$, $p < .001$, $\eta^2 = .27$. This is due to the fact that the probe-Iall difference varied across groups.

Since the inclusion of innocent group may have inflated the interaction between groups and stimulus types, we additionally conducted a 4×2 analysis with only the simple guilty, and the 2-, the 4-, and the 6-countermeasure groups as the levels of the between-subject variable. The result largely replicated the previous analysis: regarding main effect; only the group effect was significant ($F(4,47) = 4.029$, $p < .02$, $\eta^2 = .205$). *Post hoc* Tukey test showed that the RT in the 6-countermeasure group was significantly longer than that in the simple guilty group ($p < .03$), as well

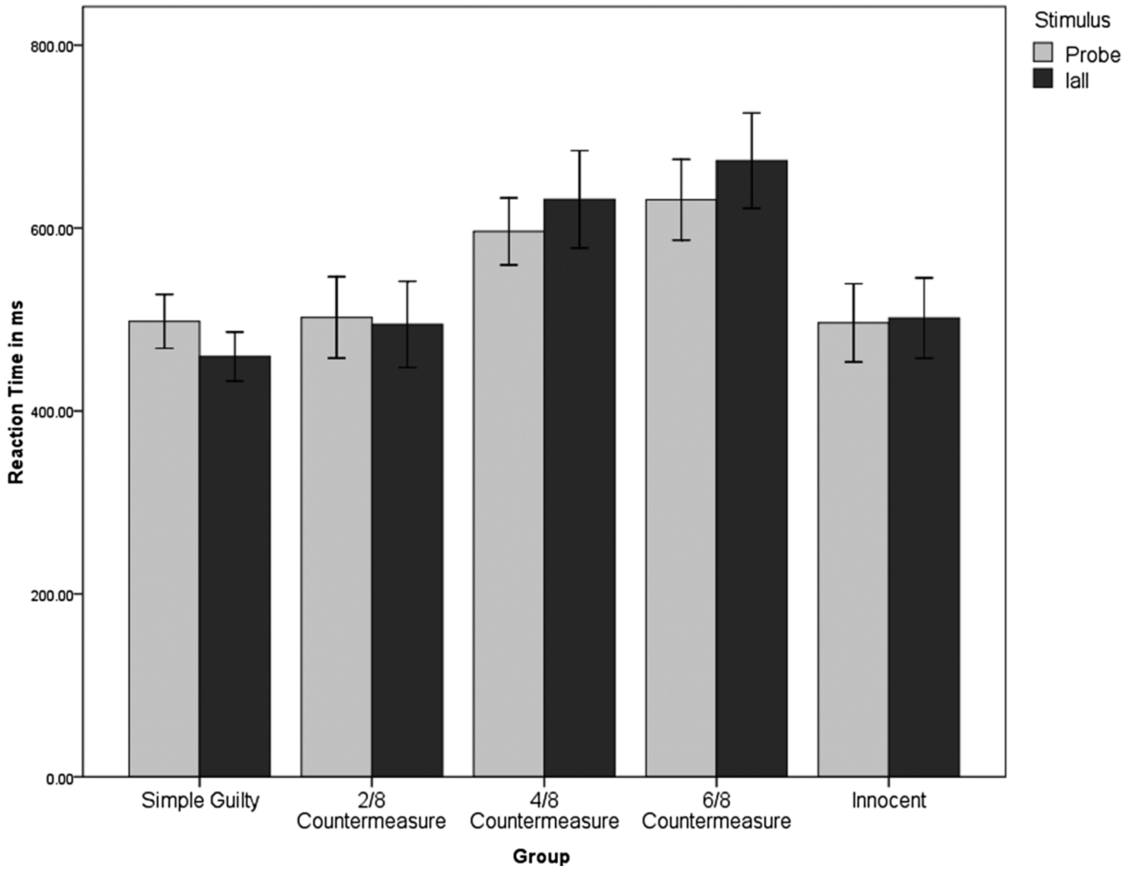

Figure 3. Reaction time (in milliseconds) of the first "I saw it" response for probe and Iall in the simple guilty group, in the 2-, 4-, and 6-countermeasure groups, as well as in the innocent group. The error bar stands for one standard error (*SE*).

as in the 2-countermeasure group at a nearly significant level ($p < .06$). The interaction between groups and stimulus types was again significant ($F(3,47) = 6.084$, $p < .001$, $\eta^2 = .28$), suggesting that the interaction in the 5 × 2 ANOVA did not dependent on the innocent condition. Follow-up analysis showed that the probe-Iall RT difference was significant in the simple guilty group ($t(11) = 3.583$, $p < .01$) with probe RT greater than Iall RT, but not in the 2- and 4-countermeasure groups ($p > .1$). Importantly, the probe-Iall RT difference was reversed in the 6-countermeaure group ($t(13) = -3.193$, $p < .01$).

These behavioral differences between countermeasure groups (see Table 1) suggested that the RT tended to be slower as the number of countermeasures increases. To test this hypothesis in another way, linear regression analyses were run using the number of countermeasures as an independent measure and the RT of probe and Iall as two separate dependent measures. Results showed that the number of countermeasures used can significantly predict the RT of both the probe ($\beta = 0.34$, $t = 2.198$, $p < .05$) and the Iall ($\beta = 0.384$, $t = 2.531$, $p < .05$). Thus, the more countermeasures participants use, the slower they will be, even when participants are using the simultaneous countermeasures.

We next focused on the analysis of P300. From Figures 4a and b, it appears that the probe-Iall amplitude difference was greatest in the simple guilty group and was smallest in the innocent group. We performed a 5 × 2 mixed ANOVA using groups as the between-subject variable (simple guilty vs. 2- vs. 4- vs. 6-countermeasure vs. innocent group) and stimulus type as the within-subject variable (probe vs. Iall). Results revealed a clear effect of stimulus type: $F(1,58) = 159.038$, $p < .001$, $\eta^2 = .733$, as evidenced by the expectedly larger p-p P300 to probe than to Iall in simple guilty and countermeasure conditions. The group effect was also significant: $F(4,58) = 6.698$, $p < .001$, $\eta^2 = .316$. *Post hoc* Tukey tests revealed that the overall P300 (probe and Iall combined) in the simple guilty group was larger than that in the 6-countermeasure group ($p < .02$) and in the innocent group ($p < .001$). Moreover, the P300 in the 2-countermeasure group and in the 4-countermeasure group was larger than that in the innocent group ($p < .01$, $p < .05$, respectively). However, the P300 in the 6-countermeasure group was not different from that in the innocent group ($p > .1$). Finally, the P300s in the 2-, 4-, and 6-countermeasure groups were not different from one anther ($p > .05$).

The group × stimulus type interaction was also significant: $F(4, 8) = 12.969$, $p < .001$, $\eta^2 = .472$. The interaction was probably due to the fact that although the difference between P300 to probe and Iall existed in the simple guilty and the three countermeasure groups (all $p < .001$), this difference was not significant in the innocent group ($p > .3$).

As with the RT analysis, we excluded the innocent condition and conducted a 2 × 4 mixed ANOVA using simple guilty, 2-, 4-, and 6-countermeasure groups as the between-subject variable and stimulus type as the within-subject variable (probe vs. Iall). Results showed that the main effect of stimulus type was highly significant ($F(1,47) = 175.379$, $p < .001$, $\eta^2 = .789$), due to the larger P300 for probe than for Iall. The group effect was also significant

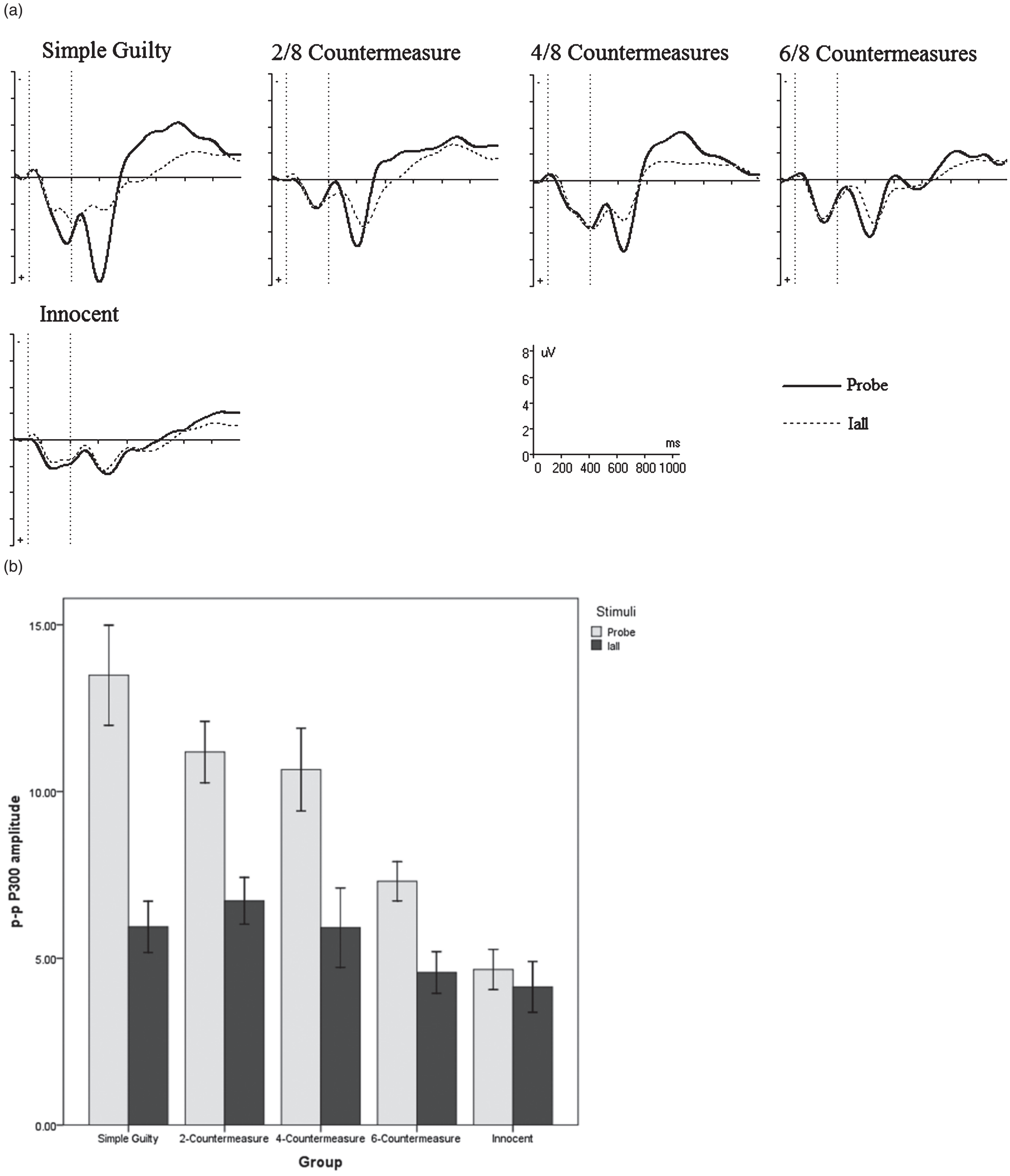

Figure 4. (A) Grand average P300 at Pz for probe and irrelevant (Iall) in the simple guilty, 2-, 4-, and 6-countermeasure groups, and the innocent groups. The first tick marker indicates the onset of the stimulus presentation; the second tick marker indicates the 300 ms after the stimulus presentation. (B) Bar graph of the same data. The error bar stands for one standard error (*SE*).

($F(3,47) = 3.514$, $p < .05$, $\eta^2 = .183$). A Tukey *post hoc* test showed that the overall P300 in the simple guilty group was larger than that in the 6-countermeasure group ($p < .03$), but no other difference was significant. Notably, the group × stimulus type interaction was again significant ($F(3,47) = 7.356$, $p < .001$, $\eta^2 = .32$). Follow-up tests revealed that whereas the group effect was significant for probe P300 ($F(3,47) = 5.754$, $p < .01$), there was no effect for Iall P300 ($F < 1$, $p > .3$).

Table 1. *Mean of Reaction Time (RTs) and Standard Deviation (SD) in Milliseconds of Probe, Iall, and Probe and Iall for Participants in Each of the Five Conditions*

	Probe		Iall		Probe and Iall	
Groups	Mean	*SD*	Mean	*SD*	Mean	*SD*
Simple guilty	498.09	102.26	459.59	92.90	463.43	93.21
2-Countermeasure	502.27	160.29	494.73	169.78	495.48	168.68
4-Countermeasure	596.37	126.88	633.87	198.87	627.54	176.58
6-Countermeasure	630.99	164.97	654.11	192.41	673.93	194.90
Innocent	496.42	148.28	501.64	152.30	502.04	148.75

The Four-Irrelevant Versus Eight-Irrelevant Protocol Comparison

One of the main motivations for the current study was to better identify simultaneous countermeasure users, which were undetectable when only four irrelevant stimuli were used in a previous study (Sokolovsky et al., 2011). It may be therefore informative to conduct a direct comparison between the current dataset and the dataset from the Sokolovsky et al. study. Specifically, the countermeasure users in the Sokolovsky et al. study were asked to execute simultaneous countermeasures to two out of four irrelevant stimuli (2/4 countermeasure group). Thus, two analyses were conducted.

First, when controlling for the number of countermeasures used in these two studies, RT from the 2/8 and the 2/4 countermeasure groups were analyzed. Specifically, we ran a 2×2 between-subject ANOVA on the RTs to probe and Iall. The first variable was the group (simple guilty vs. countermeasure group); the second variable was the irrelevant number in each study (four vs. eight irrelevant stimuli). This analysis did not find any significant results (all $F < 1$, $p > .3$). However, this analysis had a proportion of countered-irrelevants confounded: 50% (2/4) vs. 25% (2/8).

Thus, we controlled for the proportion of countermeasures executed in these two studies: the RT from 4/8 and the 2/4 countermeasure groups were included in the next analysis. A 2 (four vs. eight irrelevant stimuli) $\times$ 2 (simple guilty vs. countermeasure groups) between-subject ANOVA on the RT of probe and Iall was conducted. Result showed that the number of irrelevants had a significant effect on the RT of probe ($F(1,44) = 7.112$, $p < .02$, $\eta^2 = .139$) and Iall ($F(1,44) = 8.22$, $p < .01$, $\eta^2 = .157$), suggesting that increasing the number of irrelevants does lead to a longer RT. Moreover, the countermeasure group had a significantly longer RT than that in the simple guilty group for Iall ($F(1,44) = 6.236$, $p < .02$, $\eta^2 = .124$; see Figure 5a) but not for probe ($p > .1$; see Figure 5b). Finally, an interaction between number of irrelevants and group was significant on Iall ($F(1,44) = 5.423$, $p < .03$, $\eta^2 = .11$) but not on probe ($p > .05$). A follow-up analysis showed that when the number of irrelevants was four, there was no difference between simple guilty and the countermeasure group ($t(22) = .154$, $p > .8$), whereas when the number of irrelevants was increased to eight, the difference between simple guilty and the countermeasure group was highly significant ($t(22) = 2.886$, $p < .01$). This pattern of results again supported our hypothesis that while participants were executing a fixed proportion of countermeasures, increasing the number of irrelevants will make the use of countermeasures detectable.

Individual Detection Rates and Receiver Operating Characteristic (ROC) Analysis

Detection rate based on probe and Iall as well as the Grier A' value within each group is presented in Table 2. The A' values were calculated separately for simple guilty, 2-countermeasure, 4-countermeasure, and 6-countermeasure groups using the false positive rate from the innocent group. Overall, the correct classification rate was 87.3% (55/63) across five conditions. For calculating the Grier A' value, we used the false positive rate (8%) in the innocent condition and hit rate in each of the other four groups. These values suggested that the complex trial protocol remained effective in simple guilty ($A' = .98$) and various simultaneous countermeasure conditions ($A' = .89 \sim .96$).

Screening Countermeasure Users Based on the Individual RT

When confronted with a negative result based on P300, one should attempt to determine whether the subject is innocent or if he is a countermeasure user who defeats the test. An abnormally long RT could indicate countermeasure use, as shown in previous studies (Rosenfeld et al., 2008; Winograd & Rosenfeld, 2011). Yet, the recently found simultaneous countermeasure made this indicator ineffective when only four irrelevant items were used (Sokolovsky et al., 2011). Therefore, we increased the number of irrelevant stimuli used in the current study to eight in order to better detect the simultaneous countermeasures. The elevation of RT in the countermeasure conditions, as we found here, is encouraging. To best screen the countermeasure users from innocent subjects without increasing the false positive rate, we adopted a cutoff which was the innocent participants' mean RT plus two standard deviations; (participants' RTs and *SD*s across the five conditions are presented in Table 1). In a field application where the probe is unknown to the authority because the crime has not yet been committed (as in Meixner & Rosenfeld, 2011), we must examine the RT across all stimuli including probe and Iall, which in the present study yields a cutoff of about 800 ms. However, in the crime situation where the probe is known to the authority (as in Winograd & Rosenfeld, 2011), we can simply utilize only the RT of Iall. The cutoff in that case in the present dataset would actually be a similar 806 ms. Overall, of all the negative outcomes, one additional participant in the 6-countermeasure was screened as a countermeasure user. This slightly improved the detection rate of the 6-countermeasure group from 71% in Table 2 to 79%, and the corresponding Grier A' was increased to .91.

Discussion

The present data show that by increasing the number of irrelevant items to eight, the complex trial protocol is often able to detect simultaneous countermeasures, which were previously reported (with only four irrelevants) to make the RT index of countermeasure use completely ineffective (Sokolovsky et al., 2011). The RT's ability to indicate countermeasure use was supported by the fact that the RT pattern of the probe-Iall difference was reversed in the countermeasure groups compared with the simple guilty group. In the simple guilty group, the classic concealed information effect replicated previous RT-Concealed Information Test studies: the RT of probe was slower than the RT of Iall (Seymour & Kerlin, 2008; Seymour et al., 2000), even with no explicit probe-irrelevant discrimination task in the complex trial protocol (as in Rosenfeld, Tang, Meixner, Winograd, & Labkovsky, 2009; Winograd &

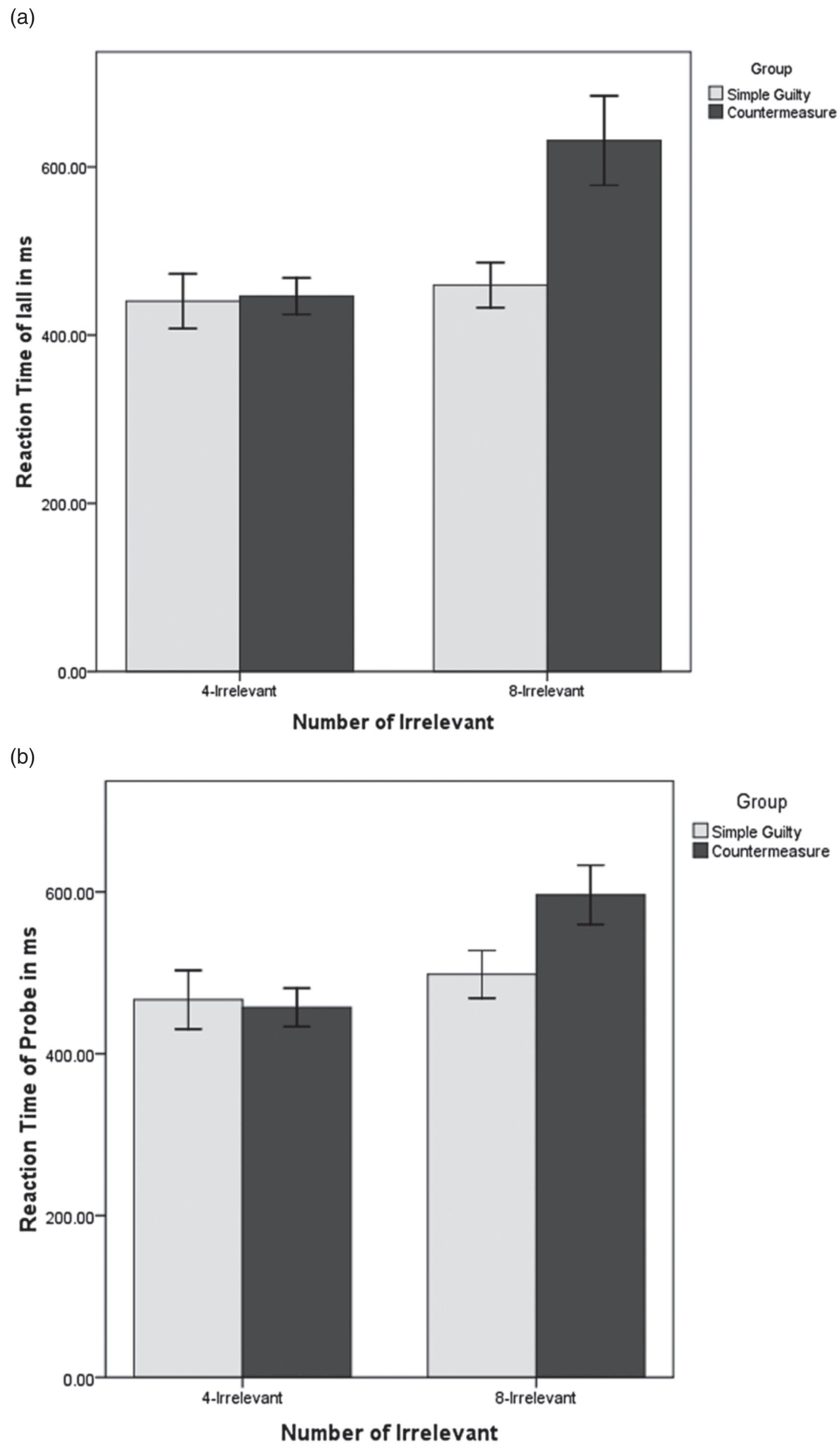

Figure 5. (A) Reaction time (in milliseconds) of the first “I saw it” response for Iall of simple guilty and countermeasure condition, controlling for the proportion of countermeasure used (2/4 vs. 4/8), from Sokolovsky et al.’s 4-irrelevant protocol and the current study’s 8-irrelevant protocol. (B) Reaction time for the probe. The error bar stands for one standard error (*SE*).

Rosenfeld, 2011). However, in the countermeasure condition, the RT of Iall was significantly longer than the RT of the probe (as in Winograd & Rosenfeld, 2011). This abnormal elevation of RT could possibly be used as a countermeasure indicator in field use. In detail, the RT of the countered irrelevant was larger than the RT to the probe and to the noncountered irrelevant. Moreover, the RT increased as the number of countered stimuli increased (see Table 1 and Figure 3). Given the previous finding that there was no RT difference between simple guilty and simultaneous countermeasure conditions in a complex trial protocol with only four irrelevant items (Sokolovsky et al., 2011), we ascribe the present findings in part to the elevation of the number of irrelevant items from four to

Table 2. *Individual Detection Rate in Each Condition and the Corresponding Grier* A*'s Value*

Group	Correct detections	Percentage	Grier A'
Simple guilty	12/12	100%	0.98
2-Countermeasure	12/13	92%	0.96
4-Countermeasure	10/12	83%	0.93
6-Countermeasure	10/14	71%	0.89
Innocent	11/12	92%	—

eight. This finding, as we expected, is probably because participants find it more cognitively demanding to select and to execute countermeasures to several among the eight irrelevants, as evidenced by the correspondingly increased RT. However, we acknowledge that the increase in number of irrelevants to eight is not a complete solution, inasmuch as the 2-countermeasure group's RTs were not affected by the manipulation. We suspect that greater increases in irrelevant numbers (to 10 or 12) might have the desired effect, but this is an empirical question.

Moreover, it must be acknowledged that the present dataset does not prove the causal connection between number of irrelevants used and RT to countered items: Even when comparing consistent proportions of countered irrelevants in the present study (with four countered of eight irrelevants) versus the previous study (of two countered of four irrelevants; Sokolovsky et al., 2011), and finding that the present but not the previous study shows elevated irrelevant RTs, one cannot safely attribute the elevation solely to increasing the irrelevant number, since the number of countered irrelevants is also increasing in a confounded manner with the irrelevant number. The study in which these two likely contributors to task demand (and RT)—irrelevant number and countermeasure number (and the way in which they may interact)—are teased apart has not yet been done. Our aim was only to show that increasing the number of irrelevants from four to eight would allow RT-based detection of countermeasure use, which it did when four or more countermeasures were used. As acknowledged above, use of two of eight simultaneous countermeasures was not detected with RT, but from an applied perspective, this is not so great a loss since, as discussed below, use of two countermeasures does not much impact P300 detection of concealed information. (Likewise, when only one of four irrelevant items is countered, P300 detects concealed information 92% of the time; Rosenfeld, 2011).

Our present hypothesis is that the demand effect of irrelevant number becomes more important as number of countered items increases. It seems reasonable to conclude that the number of irrelevants used should indeed be increased in future studies for at least two reasons: (1) Use of only four irrelevants removes the ability to detect countermeasure use with two countered irrelevants (Sokolovsky et al., 2011); and (2) Obviously, increasing the number of irrelevants reduces probe probability, which should lead to larger probe P300s (Fabiani, Gratton, & Coles, 2000). Moreover, for the subject who is not using countermeasures, the implicit task of probe recognition is not made more demanding with the addition of more irrelevant items, so that adding more irrelevants will only increase probe P300. Given that a subject is confronted with a larger number of irrelevants (8–12), it is clear that his cognitive task must increase if he chooses to increase the number (i.e., proportion) of irrelevants countered. This is because the current, and we believe optimal, countermeasure strategy from both a theoretical as well as empirical perspective is to execute a different, unique (covert) countermeasure response to each irrelevant. This is the strategy that converts the irrelevant into a P300-generating target. Executing random countermeasure responses to various irrelevants deprives them of their distinct meaningfulness, which is a P300-generating target attribute. Indeed, we have shown that performing random responses to irrelevant stimuli fails to impact probe P300 amplitude in subjects detected with a 100% hit rate (probe vs. Iall); neither was RT impacted in this group (Meixner, Haynes, Winograd, & Rosenfeld, 2009).

In addition to the RT finding, the P300-based complex trial protocol here remains effective in detecting concealed information in various countermeasures conditions. Specifically, the proportion of the to-be-countered irrelevant items was systematically varied from .25 to .5 to .75 here. The goal of this manipulation was to examine the complex trial protocol's resistance to various numbers of simultaneous countermeasures. The fact that the countered irrelevants elicited a larger P300 than noncountered irrelevants suggests that participants indeed attempted to beat the test. Moreover, the amplitude of probe P300 decreased and the RT increased as the countered irrelevant number increased. This suggests that higher task demand systematically decreased the P300 to the probe (as earlier reported by Kramer et al., 1987). Similarly, the overall P300 across Iall and probe in the 6-countermeasure group was significantly smaller than that in the 2-countermeasure group. This is further evidence that the P300 decreases as task demand increases. However, despite the fact that the P300 of the probe was systematically influenced by countermeasure use, the difference between the probe and irrelevants remained robust. This probe-Iall difference supported the complex trial protocol's effectiveness with eight irrelevants in resisting even simultaneous countermeasures, as reflected in individual detection results.

First, the classification in the simple guilty and innocent groups remains accurate (100% hit rate with 8% false positive), supporting the P300 complex trial protocol's promise in future memory detection use. Second, the P300-based test successfully identified 82% (see Table 2) of all countermeasure users as guilty before countermeasure screening. Although the detection rate is lower in the most challenging 6-countermeasure condition compared with the 2- and 4-countermeasure conditions (71% vs. 83% and 92%, respectively), this is still superior to the 50% or lower reported in previous P300 studies (Mertens & Allen, 2008; Rosenfeld et al., 2004) not using the complex trial protocol. It also appears that the most effective countermeasure strategy to defeat the test would be conducting countermeasures to a higher proportion of irrelevant items (e.g., .75 here, but not to all irrelevants, as suggested by Meixner & Rosenfeld, 2010). However, although using more countermeasures is associated with reduced P300 and reduced detection rate, it should be also noted that the RT is also increased correspondingly, which increases the probability of countermeasure use detection. Moreover, it is clear that the increased RT associated with a higher proportion of countered irrelevants suggests that many participants may find it prohibitively difficult to counter six out of eight irrelevants.

Also, it should be recalled that all participants in the countermeasure groups had learned and practiced countermeasures prior to the complex trial protocol, whereas in the field, the countermeasures would have to be associated with the irrelevants on the spot. With six of eight irrelevants, this could be a daunting task.

The field is where we ultimately want to apply the new complex trial protocol, and one might be concerned that hometown names do not comprise a typical category encountered in field forensic CIT tests. Actually, one could encounter this type of information in tests of individuals feigning memory deficit subsequent to a mild head injury. However, this information type was used here simply for illustrative purposes, and because it is easy to obtain from school records. While the experimenters knew the information relevant for each subject at the point when the stimulus lists were prepared, by the time a participant was tested, the operator no longer knew the information and, in any case, the subject was to conceal the information so that the P300-based test, not the experimenter, could not detect it.

To conclude, the present study showed that the complex trial protocol with eight irrelevant items could resist even simultaneous countermeasures in various conditions. Given that countermeasures pose a serious threat to the P300-based memory detection, the current result is promising. It may also be encouraging for other neuroscience-based memory detection methods such as fMRI, which was also recently found to be vulnerable to countermeasures (e.g., Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011). Future study is warranted to address this issue.

References

Allen, J., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology*, *29*, 504–522. doi: 10.1111/j.1469-8986.1992.tb02024.x

Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge test technique: Effects of mental countermeasures. *Journal of Applied Psychology*, *81*, 273–281. doi: 10.1037/0021-9010.81.3.273

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). San Diego, CA: Academic Press.

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

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. Berntsen (Eds.), *Handbook of Psychophysiology* (pp. 85–119). New York, NY: Cambridge University Press.

Ganis, G., Rosenfeld, J. P., Meixner, J. B., Kievit, R. A., & Schendan, H. (2011). Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage*, *55*, 312–319. doi: 10.1016/j.neuroimage.2010.11.025

Grier, J. B. (1971). Non-parametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429. doi: 10.1037/h0031246

Honts, C. R., & Amato, S. L. (2002). Countermeasures. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 251–264). San Diego, CA: Academic Press.

Honts, C. R., Amato, S. L., & Gordon, A. K. (2001). Effects of spontaneous countermeasures used against the comparison question test. *Polygraph*, *30*, 1–10.

Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, *33*, 84–92. doi: 10.1111/j.1469-8986.1996.tb02111.x

Kramer, A. F., Sirevaag, E. J., & Brauner, R. (1987). A psychological assessment of operator workload during simulated fight missions. *Human Factors*, *29*, 145–160.

Lykken, D. T. (1998). *A tremor in the blood*. Reading, MA: Perseus Books.

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

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

Meixner, J. B., & Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300-based Concealed Information Test. *Psychophysiology*, *47*, 57–65. doi: 10.1111/j.1469-8986.2009.00883.x

Meixner, J. B., Haynes, A., Winograd, M. R., & 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 and Biofeedback*, *34*, 209–221. doi: 10.1007/s10484-009-9091-4

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

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, *38*, 319–335. doi: 10.1111/j.1469-8986.1991.tb02202.x

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

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

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

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. doi: 10.1111/j.1469-8986.2004.00158.x

Rosenfeld, J. P., Tang, M., Meixner, J. B., Winograd, M., & Labkovsky, E. (2009). The effects of asymmetric versus symmetric probability of targets following probe and irrelevant stimuli in the Complex Trial Protocol (CTP) for detection of concealed information with P300. *Physiology & Behavior*, *98*, 10–16. doi: 10.1016/j.physbeh.2009.03.030

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

Seymour, T. L., & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology*, *25*, 475–490. doi: 10.1002/acp.1375

Seymour, T. L., Seifert, C. M., Mosmann, A. M., & Shafto, M. G. (2000). Using response time measures to asses "guilty knowledge." *Journal of Applied Psychology*, *85*, 30–37. doi: 10.1037/0021-9010.85.1.30

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). The case for peak-to-peak measurement of P300 recorded at .3 Hz high pass filter settings in detection of deception. *International Journal of Psychophysiology*, *40*, 173–180. doi: 10.1016/S0167-8760(00)00154-9

Sokolovsky, A., Rothenberg, J., Meixner, J., & Rosenfeld, J. P. (2011). Effect of sequential versus simultaneous stimulus acknowledgement and countermeasure responses on P300 based detection of deception in the Complex Trial Protocol. *International Journal of Psychophysiology*, *81*, 60–63. doi: 10.1016/j.ijpsycho.2011.03.008

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, *26*, 208–221. doi: 10.1111/j.1469-8986.1989.tb03159.x

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

(Received March 10, 2011; Accepted July 14, 2011)