Superiority of Pictorial Versus Verbal Presentation and Initial Exposure in the P300-Based, Complex Trial Protocol for Concealed Memory Detection
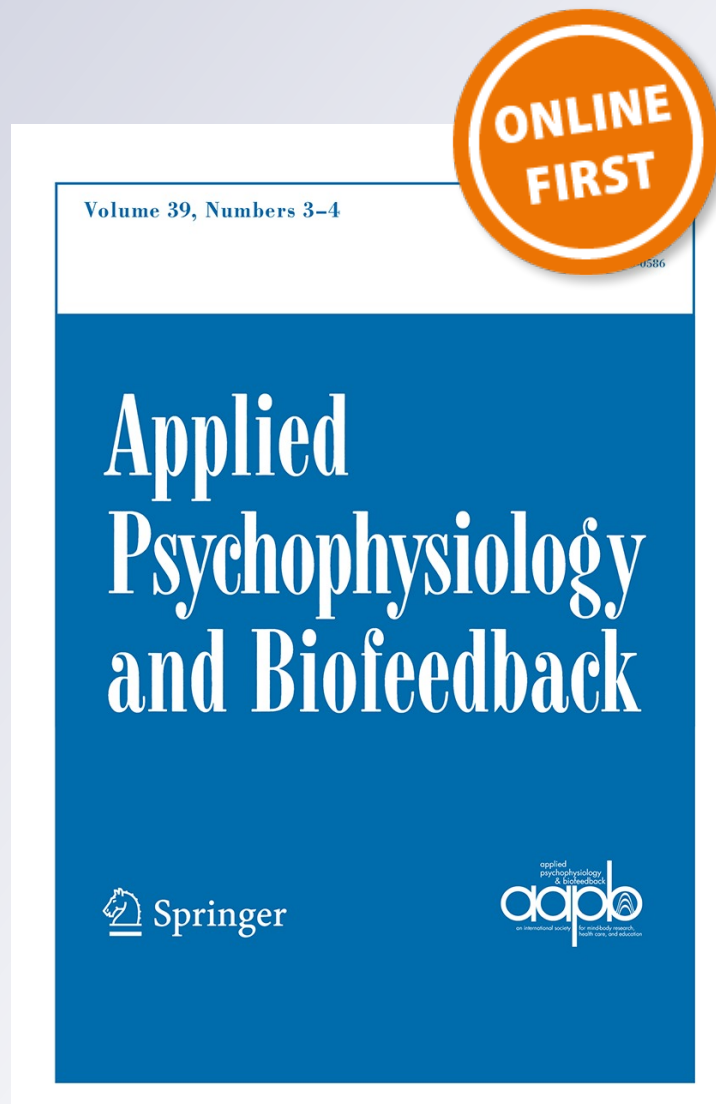
J. Peter Rosenfeld, Anne Ward, Michelle Thai & Elena Labkovsky

ONLINE FIRST

Applied Psychophysiology and Biofeedback

Springer

aapb
applied psychophysiology & biofeedback
an international society for mindbody research, health care, and education

Springer

Springer

Superiority of Pictorial Versus Verbal Presentation and Initial Exposure in the P300-Based, Complex Trial Protocol for Concealed Memory Detection

J. Peter Rosenfeld[1] · Anne Ward[1] · Michelle Thai[1] · Elena Labkovsky[1]

Abstract Two mock guilty groups had either pictorial or verbal initial exposure to crime items (probes) on which they were told they would later be tested. Then each subject was tested in two sessions on two successive days with both verbal and pictorial presentation, one test modality per session/day. The three dependent variables analyzed were three different estimates of the same basic measurement: the difference between P300s evoked by key (probe) and irrelevant stimuli. All three indexes were significantly increased more by both initial pictorial exposure, as well as by pictorial presentation modality, than by verbal exposure and presentation. We saw no main effect of exposure–presentation modality congruence, as congruence interacted with exposure: The largest probe–irrelevant differences were evoked by congruent pictorial exposure and presentation modality, and the smallest by congruent verbal exposure and presentation modality.

Keywords P300 · Guilty knowledge tests · Concealed information tests · Lie detection · Pictorial superiority · Memory detection

Introduction

A suspect's guilt or innocence may be sometimes *inferred* in the *concealed information test* (CIT), also known as the *guilty knowledge test* (GKT, Lykken 1959). The CIT or

GKT assumes that a guilty person possesses critical information known only to the police, the victims and persons involved in committing the crime. The CIT or GKT presents a series of information-containing stimulus items to a subject, one at a time. One or more of these items may be directly relevant to the crime under investigation and these are called critical or *probe* items. The other, more frequently presented items are irrelevant to the crime and they are called *irrelevant*. The CIT/GKT detects whether the suspect recognizes (is knowledgeable of) the critical probe information. In other words, CITs are used to learn if an individual recognizes one or more crime-related items of information. It is assumed that only guilty (knowledgeable) persons (but not innocent, un-knowledgeable persons) can recognize such information and thus involuntarily respond to it with an enhanced physiological response indicative of recognition.

To optimize concealed information detection tests (CITs), which have evolved over the years using increasingly sophisticated methodologies (Rosenfeld et al. 2012), it is nevertheless important to pin down fundamental testing parameters which may vary across protocols using differing dependent measures; autonomic, imaging, and electroencephalographic. One of the most basic of these fundamental parameters involves the modality chosen for presentation of CIT questions.

In the field, questions are typically put to suspects acoustically, although more recently, and especially in the laboratory use of event-related potentials (ERPs) and imaging, questions are usually presented verbally on a display screen (e.g., Rosenfeld 2011). Even when in one CIT, multiple items of information are probed, they are nevertheless presented one at a time. Thus a crime related item such as "356 MAGNUM" can be presented verbally, just as shown within the preceding quotation marks, or it

✉ J. Peter Rosenfeld
jp-rosenfeld@northwestern.edu

[1] Department of Psychology, Institute for Neuroscience, Northwestern University, Evanston, IL, USA

can be presented in pictorial form. Indeed, Labkovsky and Rosenfeld (2014) recently reported a new P300-based CIT in which mock crime-related stimuli were presented verbally or pictorially, with both presentation modalities able to evoke P300 waves in knowledgeable subjects. However, the design of this study obviated an un-confounded comparison of the effects of verbal versus pictorial presentation mode. This is because the pictorially presented items were always office or personal-type items, whereas the verbally presented items were always names. Thus the verbal–pictorial comparisons were confounded with the nature of the stimulus, names versus office/personal items. Therefore, one present major aim is to directly test (in an un-confounded manner) whether pictorial versus verbal presentation of stimuli might lead to larger P300s and superior accuracy in detection of concealed information.

There have been two such studies of which we are aware: one was that of Ambach et al. (2010), who reported no difference between verbally versus pictorially evoked P300s in a CIT. This study, however, used methods which we (Rosenfeld 2011) have found not well suited to P300 usage in detection of concealed information: First, they used a standard baseline-to-peak (b–p) measurement of P300 amplitude which we and others (Meijer et al. 2007; Soskins et al. 2001) have found at least 25 % less accurate in P300-based CITs than the peak–peak (p–p) measure used here and described below in the methods section. Second, they used a *multiple probe protocol* with several different averaged probes (guilty knowledge items) in one block, whereas we have shown that this protocol is less sensitive and more demanding than a protocol using one probe per block (Rosenfeld et al. 2004, 2007. See also Lui and Rosenfeld 2008). Third, most importantly, they measured P300 from 400 to 1000 ms, whereas it is evident from the average ERPs in their Fig. 3—in which pictorially evoked P300s appear grossly larger than verbally evoked P300—that the P300 peak for both visually and pictorially evoked P300s seems to fall at about 350 ms. Thus it started even earlier than 300 ms. They likely missed measurement of the true P300 peak in many subjects.

The other such study was by Cutmore et al. (2009). They compared presentation modalities for pictures, faces, and words. Their picture–word comparison was similar in intention to that of the present similar comparison, however their visual and pictorial stimuli were of different area sizes, and neither the actual dimensions nor retinal angles were specified, nor were there illustrations of their stimuli presented. It was therefore difficult to interpret the reported findings of superior picture presentation modality (which is actually similar to what we report below). Neither Cutmore et al. (2009) nor Ambach et al. (2010) made any attempt (as we do here) to manipulate initial exposure modality, so that exposure–presentation congruence effects could not be studied. (This initial exposure effect is discussed below in the paragraph after next.)

Seymour and Kerlin (2008) also found no pictorial versus verbal presentation difference, using reaction time (RT) as the dependent measure in their CIT. Neither did CIT studies based on the electrodermal response (EDR, Ben-Shakhar et al. 1996; see reviews in Ambach et al. 2010 and in Seymour and Kerlin 2008). Indeed, several studies from Ben-Shakhar's lab reported an advantage for verbal over pictorial presentation in EDR studies of the orienting response commonly assumed to underlie CIT differential responding (see Ben-Shakhar and Gati 2003, and citations therein to the same authors). The ANS-based CIT is at least in part memory-based, and with both recall and recognition procedures in widely disparate fields, pictorial memory has been long reported as superior to verbal memory of matched items (Kirkpatrick 1894; Shepard 1967; Wells 1972; Gehring et al. 1976; Park et al. 1983; Childers et al. 1985; Stenberg 2006). Since most researchers agree that the P300 CIT is also at least partly recognition-based (see Rosenfeld 2011 for review), we suggest it is important for the field to have further study of verbal versus pictorial presentation effects, especially on sensitivity and specificity of the P300-based Complex Trial Protocol (CTP) version of the CIT (Rosenfeld et al. 2008; Rosenfeld 2011). The studies of Cutmore et al. (2009) and of Ambach et al. 2010 were based on the older "3-stimulus protocol" version of the P300 CIT (described in Rosenfeld 2011), which has been found to be more vulnerable to countermeasures (CMs) than the CTP (Labkovsky and Rosenfeld 2014; Rosenfeld 2011). This seemed to us another good reason to investigate the verbal–pictorial presentation issue using the CTP, which has never been previously done.

Finally, there is the related issue of initial exposure modality of later-tested items to be considered. In the field, crime-related informational items are often initially exposed to subjects by their direct handling and pictorial visualization of crime-related items from multiple angles, providing somatic sensory as well as pictorial information (e.g., about a gun, an amount of money stolen, an object stolen, etc.). However, other crime-related items such as documents taken or photographed (e.g., in an espionage case) may be initially exposed in a principally verbal format. Thus, since the modality of initial exposure of probe items may interact with presentation modality during a CIT, we reasoned that initial exposure modality effects should also be systematically investigated in the present study.

The rather complex, related background literature on this question deals with the exposure-test presentation modality interaction concerning memory performance, and does suggest a positive effect of exposure–presentation congruence (Graf et al. 1985; Tversky 1969; Heckler and

Childers 1992; Peeck 1974; Schacter and Graf 1989; and especially Stenberg et al. 1995). The well-known "Pictorial Superiority" effect (Stenberg 2006), and its extension to ERP studies (e.g., Herron and Rugg 2003) indeed refers mainly to the *exposure* modality effect. Herron and Rugg (2003), for example, showed that remembered "old" (studied) words evoked a larger *late positive parietal potential* than "new" words in an 'old-new', study-test paradigm, and this old-new effect was larger with pictorial than verbal exposure. However, there were exceptions. In detail, all subjects initially studied items in a mixed list of both pictures and words. They were then tested in a block with both studied and novel test words, and instructed to press "yes" only to studied pictures, and in another block where they pressed "yes" only in response to studied words. As they stated in their abstract, "Relative to new items, correctly classified items studied in both target modalities elicited robust, positive-going "old/new" effects. When pictures were targets, test items corresponding to studied words also elicited large effects. By contrast, when words were targets, old/new effects were absent for the items corresponding to studied pictures." In their discussion, they conclude, "The picture superiority effect in recognition memory is diminished or reversed when, as in the present study, words are employed as retrieval cues."

However, Stenberg et al. (1995), in their fourth experiment, showed an advantage for the pictorial *testing* (presentation) modality on memory performance which overshadowed the exposure–presentation congruence effect. Herron and Rugg's (2003) enhanced *late positive parietal potential* does bear some resemblance to the P300 we use in our P300-based CITs. Indeed, Dien et al. (2004), argued that the old-new potentials are actually P300s, although as will be noted below, others have disassociated the late positive potentials associated with the old-new paradigm and P300.

Although these pictorial superiority and old-new studies might predict both presentation and exposure modality effects in P300 CIT studies, there are major differences between the paradigms used in these studies and the P300-based CIT scenario that suggest doing the present P300-based CIT study: (1) In most of the cited studies showing pictorial superiority, multiple stimuli are studied within one experiment, and during testing, multiple old and multiple new stimuli are also presented. In contrast, in the present P300 CIT only the probe or key stimulus is studied, as is often the case in the field as perpetrators focus on one or just a few relevant crime items. (2) In many pictorial superiority studies, multiple 2-s initial exposures are used for multiple stimuli, and often are also used as *test* stimulus durations. In the present study, exposure is much different (1.5 min total for one stimulus; see methods), and test presentation duration is 300 ms. (3)
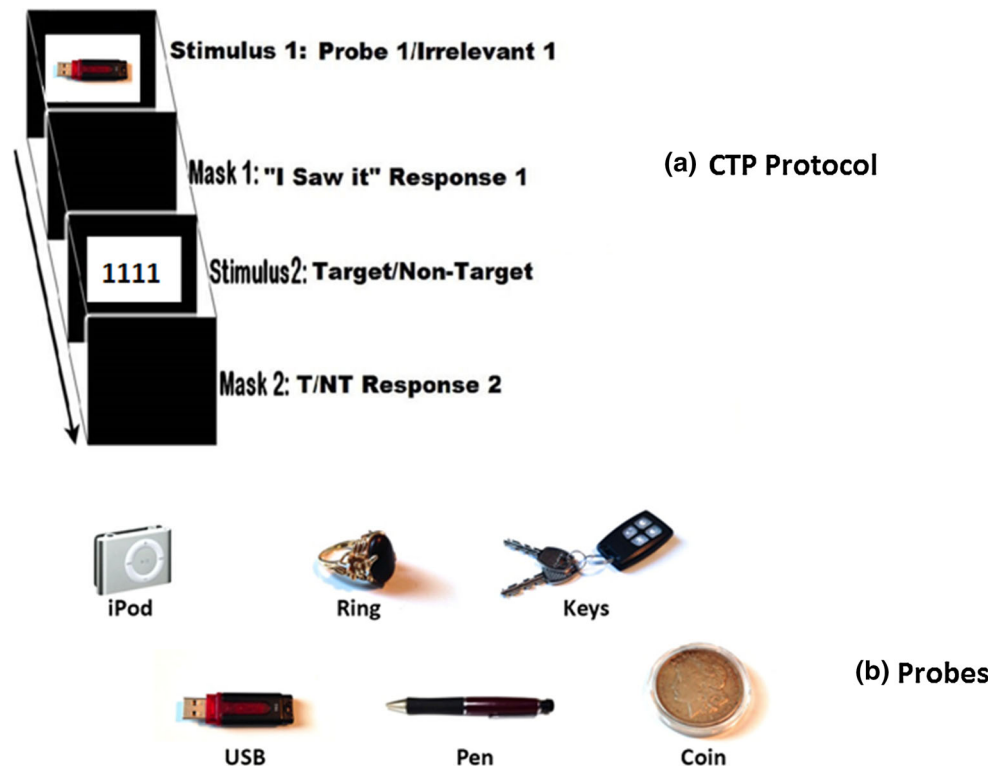
It is typically the case that a P300 CIT capitalizes on the property of P300 that the rarer the probe, the greater the P300, whereas in the pictorial superiority studies, including the ones looking at late positive "old-new" parietal potentials, the probability of the key ("old") stimuli is equated with that of "new" stimuli. Indeed, Weinberg and Hajcak (2011); Schupp et al. (2000) and others have dissociated P300 and the late positive potential. Still others describe a late positive potential (bearing some resemblance to the ERP associated with "old-new" effects) that is never related to, let alone equated with P300 (e.g., Rugg and Curran 2007). (4) Finally, most importantly, it is evident that in the predominantly behavioral studies of pictorial superiority, P300 is not the dependent variable used as is the case here; it is typically accuracy and/or RT that are measured in these *behavioral* studies.

We thus suggest that it is important to clarify *both* potential exposure and presentation modality effects, in an un-confounded manner in our countermeasure-resistant P300 CIT called the CTP. We accomplished this by assigning one subgroup of three subjects to each of one of the four counterbalance conditions (1–4) shown in Table 1. The design tested each subject on 2 days (each with two presentation modality blocks) of consistent exposure modality, with pictorial exposure in conditions 1 and 2, and verbal exposure in 3 and 4. Presentation modality (block) order in Day 1 was reversed in Day 2, necessitating that different probe stimuli be used on each block to control for habituation, requiring order counterbalance over conditions for the nuisance variable, "Probe" in Table 1. In fact, six different stimuli (in Fig. 1b) were used for day pairs across subjects, then pairs in reverse order were repeated as shown in Table 1. "Day" was another nuisance variable. The major independent effects of interest for the present study included the between subjects factor, *exposure modality*, the repeated measure variable, *presentation modality,* and their potential interaction.

Table 1 Study design

Condition	Day/expos. mode	Probe	Block 1 test	Block 2 test
I	1 Pictorial	Ring	(a) Pictorial	(i) Verbal
	2 Pictorial	Coin	(b) Verbal	(j) Pictorial
II	1 Pictorial	Keys	(c) Pictorial	(k) Verbal
	2 Pictorial	USB	(d) Verbal	(l) Pictorial
III	1 Verbal	Pen	(e) Pictorial	(m) Verbal
	2 Verbal	Ipod	(f) Verbal	(n) Pictorial
IV	1 Verbal	Coin	(g) Pictorial	(p) Verbal
	2 Verbal	Ring	(h) Verbal	(q) Pictorial

Fig. 1 a The CTP protocol event flow with a picture of a USB flash drive for Stimulus 1 (a probe or irrelevant) and the target number string "1111" for Stimulus 2. **b** The set of six stimuli used in both pictorial and verbal form

Methods

Participants

Five members of an advanced laboratory class in cognitive psychophysiology and seven of their friends voluntered as participants. (Of the total of 12, five were female, all aged 18–30). Informed Consent was obtained from all subjects and the project was approved by the Northwestern University IRB. One subject's data was corrupted, so only 11 data sets were available for final analyses. Thus, only two subjects' data sets were available for condition 4 in Table 1, versus three each for 1–3. There were six subjects available for conditions 1–2 (group 1) and five for 3 and 4 (group 2), but each subject was tested twice, yielding an adequate *df* value to find the significant effects and large effect sizes reported below.

Exposure

Participants were randomly assigned to be initially exposed to the probe ("stolen") item either pictorially or verbally. The probe-to-irrelevant ratio was 1:5. For initial exposure, participants were first shown the probe on a computer display screen for 30 s and asked to carefully examine it, as they would be later tested to see if they recognized this "stolen" item from a mock crime in which they, so they are told later, are suspected of having participated. The

stimulus was then removed and the participants were asked to visualize the stimulus and mentally recall all its details for 30 s. Participants were then exposed to the probe on a display screen a final time for another 30 s before the CIT began. The experimental design was given in the introduction and also appears in Table 1.

CTP Protocol

In the CTP used here, there were two stimuli presented in two separate parts of one trial (Fig. 1a). The first part of the trial involved presentation of either a probe or irrelevant stimulus, and the subject acknowledged stimulus perception by pressing the same left mouse button, regardless of whether the probe or an irrelevant had just been presented. We call this perception acknowledgement the "I saw it" response. Without attention-holding targets (such as those used in older P300 protocols: Rosenfeld 2011) in Part 1 of this *CTP*, attention to this first, critical trial part is maintained via expected but unpredictable testing on the identity of the first (probe or irrelevant) stimulus (see Rosenfeld 2011). That is, prior to the run, subjects were warned that there would be occasional, unpredictable pop quizzes regarding what the previous first stimulus was. Though penalties (such as loss of bonus payment) have been threatened for more than one error in these pop quizzes, in running more than 350 subjects to date, there has been <1 % attrition for this reason (Rosenfeld et al. 2013).

The second part of the complex trial (Fig. 1a) involves target or non-target presentation: After the probe or irrelevant was presented, and 1300–1800 ms (random delay) following the immediate "I saw it" response (perception acknowledgement), a target ("11111") or a non-target ("22222", "33333", etc.) is presented. The subject presses a target button or a different, non-target button to acknowledge discrimination of these two stimuli. Note that the abstract, numerical targets and non-targets have nothing to do with the crime. Targets and non-targets in this protocol were also used to help maintain attention. The targets and non-targets in this protocol were all presented in the alphanumeric/verbal modality (see Fig. 1a). The conditional probabilities of targets (or non-targets) following probes and irrelevants were equal. Overall target and non-target probability was .5. It is noted that each trial lasted 4.5 s. Since 50 probe and 250 (50 trials for each of 5) irrelevant items were presented, a session or block lasted about 30 min including breaks. After artifact removal see below), at least 30 trials of each item were experienced.

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 80uV. 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"). Eye blink artifacts were corrected with the method of Semlitsch et al. (1986). Signals were passed through Mitsar amplifiers with a 30 Hz low pass filter setting, and high pass filters set (3 db) at .16 Hz. Amplifier output was passed to the Mitsar 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, with the digital filter set 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 and others' previous studies, is the most sensitive in P300-based deception investigations (e.g., Meijer et al. 2007; Soskins et al. 2001): For P300 in this study, the algorithm searched from 300 to 700 ms for the maximally positive 100 ms segment average. This is the base–peak or b–p measure. Although we have used other (but similar) search windows in other studies, we believe it a questionable practice to choose a look window for novel studies with novel

protocols, based on search windows used in prior studies with different protocols and with different P300 latencies. Our present choice was made based on a grand average of all present subjects in all conditions, a procedure recommended by Keil et al. (2014). That is, we verified by visual inspection of each individual average that the P300 peak fell within the window bounded by 300 and 700 ms, the window that clearly contained P300 in the grand average. The midpoint of the segment defined P300 latency. (Average P300s in these studies peaked at about 500 ms on average, as seen in Fig. 2.) Then the algorithm searched from this P300 latency to 1300 ms for the maximum 100 ms negativity. The difference between the maximum positivity and negativity defines the p–p measure.

Group Analyses

ANOVA methods are used here for analysis of group/condition effects, and two effect size estimates are provided, partial eta squared (η_p^2) and "classical" eta squared (η^2). Richardson (2011) observed that although the former is preferred currently for many reasons, one cannot compare η_p^2 values for independent variables in a mixed, higher order ANOVA, for which purpose he suggested using η^2.

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 separate distributions of average probe and irrelevant 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-followed-by target in the CTP and probe-followed-by non-target in CTP) set (all single sweeps) and draws at random, with replacement, a set of n1 probe 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
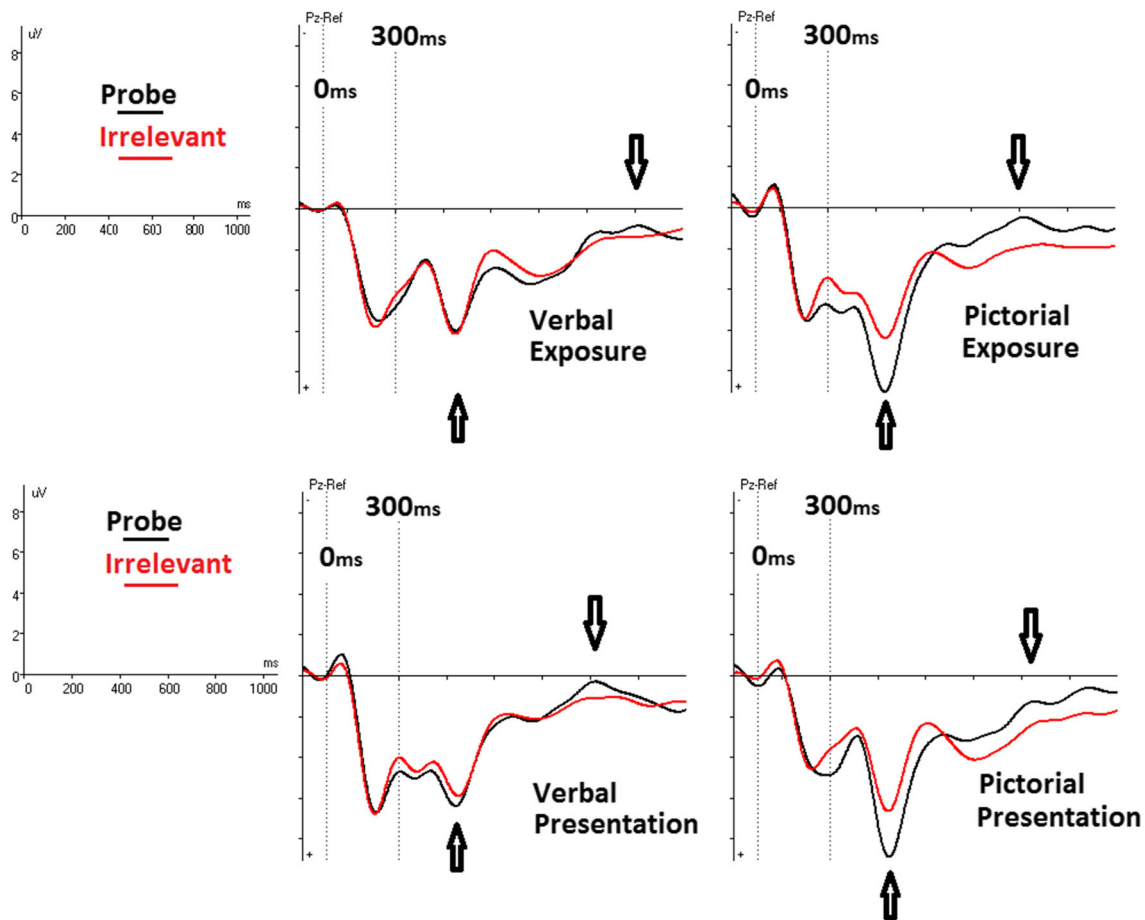
Fig. 2 Superimposed Pz probe (*black*) and irrelevant (*red*) grand average waveforms sorted by exposure modality (*top pair*) across combined presentation modalities, and presentation modality (*lower pair*) across combined exposure modalities. *Up arrows* show peaks of P300 components, and *down arrows* show peaks of negative components following P300 but within the 1300 ms look window; these peak differences yield the p–p P300 values. *Vertical faint dotted lines* at 0 and 300 ms show the stimulus onset and offset (Color figure online)

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 sweeps for that subject, and n2 is the actual number of accepted irrelevant sweeps for that subject multiplied by a fraction (about .17 on average across subjects in the present report) which randomly reduces the number of irrelevant trials to within one trial of the number of probe trials (average = 32.6 over Ss). 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. We use the mean of this distribution here as one dependent variable, as described below.

In order to state for a given subject with 90 % confidence (the criterion used in most preceding studies, e.g., Farwell and Donchin 1991; Soskins et al. 2001; Rosenfeld et al. 1991, 2004) that the (bootstrapped average) probe P300 is greater than the (bootstrapped average) irrelevant P300, and that therefore the subject recognizes the probe, at least 90 out of 100 bootstrapped P300 difference iterations must yield Probe P300 > Irrelevant P300.

Alternatively seen, if the distribution of probe–irrelevant microvolt differences can meet a normality assumption, which is often the case, we would be in effect requiring that the value of zero difference or less (a negative difference) not be >-1.29 SDs below the mean of the distribution of differences. Thus, 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 equal 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.) We emphasize that optimizing diagnostic

accuracy is not our main concern in this report. Here we focus mainly on comparison of pictorial versus verbal modalities, for both exposure and presentation: The bootstrap measures are used here mainly as dependent variables now described.

Dependent Variables

In evaluating the group effects of the critical independent variables of interest, three different dependent variables were utilized here. First, and obviously, is the Pz p–p P300 amplitude difference in microvolts (P300DF) between probe and irrelevant P300 averages, expected to be large in knowledgeable, but not in unknowledgeable subjects (a group not run here). This is a mean computed directly from the present sample of participant data. Additionally, in the intraindividual bootstrapping diagnostic proceedure we use (detailed above), means of the iterated bootstrapped average p–p P300s for probe and irrelevant items are produced in each subject for each iteration, and the mean of these sample means also estimates the *population* mean P300s for probes and irrelevants (Efron 1979). Thus our second dependent measure utilized here is the difference between such estimated population means for probe and irrelevant (BSMEAN; it has correlated >.95 with P300DF in previous studies). Finally, the *most direct measure of diagnostic accuracy* in these studies is the number of bootstrapped iterations out of the 100 performed in which the bootstrapped probe P300 (p–p) at Pz for an iteration is greater than that obtained for the bootstrapped irrelevant P300 (p–p) for the same iteration. (This is sometimes called the $P > I$ value, and is here abbreviated to BSITERS.) As noted above, for a *knowledgeable* versus *unknowledgeable* decision to be made in these studies, one usually specifies a *criterion* number of $P > I$ values that must be reached for a *knowledgeable* decision, and thus the higher the $P > I$ value, the greater the likelihood of a *knowledgeable* decision, as described above. In many recent P300 studies (Rosenfeld 2011), the criterion has been defined as .9; that is, at least 90 out of 100 iterations must yield $P > I$ for a *knowledgeable* decision, although other criteria may be used in some situations; see Rosenfeld et al. (2013). Again, we are not here concerned with diagnostic accuracy per se, but with pictorial versus verbal exposure and presentation modality comparisons.

Results

Behavioral

We collected RT data for the "I saw it" response so as to allow comparison with earlier pictorial superiority studies

(e.g., Stenberg et al. 1995) based on RT. However since this "I saw it" response is simply a perception acknowledgement, and does not involve much cognitive effort, we did not expect any interesting RT outcomes. Indeed comparing pictorial and verbal probe RTs over both days of testing, yielded all $p > .3$. The same was true for irrelevant RTs.

ERP Results, Qualitative

Figure 2 illustrates one sorting of four grand average Pz ERP superimpositions (probe black, irrelevant red). The figures in the top two panels show the effects of exposure modality (verbal left, pictorial right) averaged in each case across both presentation modalities. The figures in the lower two panels show the effects of presentation modality (verbal left, pictorial right) averaged in each case across both exposure modalities. There appear to be greater probe–irrelevant differences in the separate cases of both pictorial exposure and presentation, than in the separate cases of verbal exposure and presentation. Figure 3 shows the probe–irrelevant differences sorted by the four possible different exposure–presentation modality combinations: These are (1) pictorial exposure and pictorial presentation (hereafter PP, lower right in Fig. 3), (2) pictorial exposure and verbal presentation (PV, lower left in Fig. 3), (3) verbal exposure and presentation (VV, upper left in Fig. 3) and (4) verbal exposure with pictorial presentation (VP, upper right in Fig. 3). To the eye, it appears that for probe–irrelevant p–p amplitude differences at Pz, $PP > PV > VP > VV$, as might be expected from Fig. 2. This progression suggests main effects of both presentation and exposure modality. Line graphs of computed Pz p–p P300 amplitudes (P300DF) in microvolts, as well as numbers of $P > I$ bootstrap iterations (out of 100) for Pz p–p P300s (BSITERS), are shown in Fig. 4 as function of (presentation modality, exposure modality, and "exposure–presentation congruence"; PP, VV are congruent, PV and VP are incongruent). As the labels on the data points show, Fig. 4 simply plots the same data in differing ways in three rows so as to illustrate the differing relationships seen in various plots: The top two panels suggest the effects of presentation modality (pictorial > verbal), especially for the verbal exposure condition, but also suggest even larger exposure modality effects (pictorial > verbal), which are perhaps better visualized in the middle two panels. The lower panels clearly suggest an interaction of congruence and exposure modality, mathematically equivalent to the exposure by presentation interaction suggested in the top row.

   It is evident that within each row of two panels in Fig. 4, the left and right members look alike, as expected since both are correlated dependent measures: the greater the probe–irrelevant differences, the greater the expected numbers of $P > I$ iterations. Figure 4 confirms the
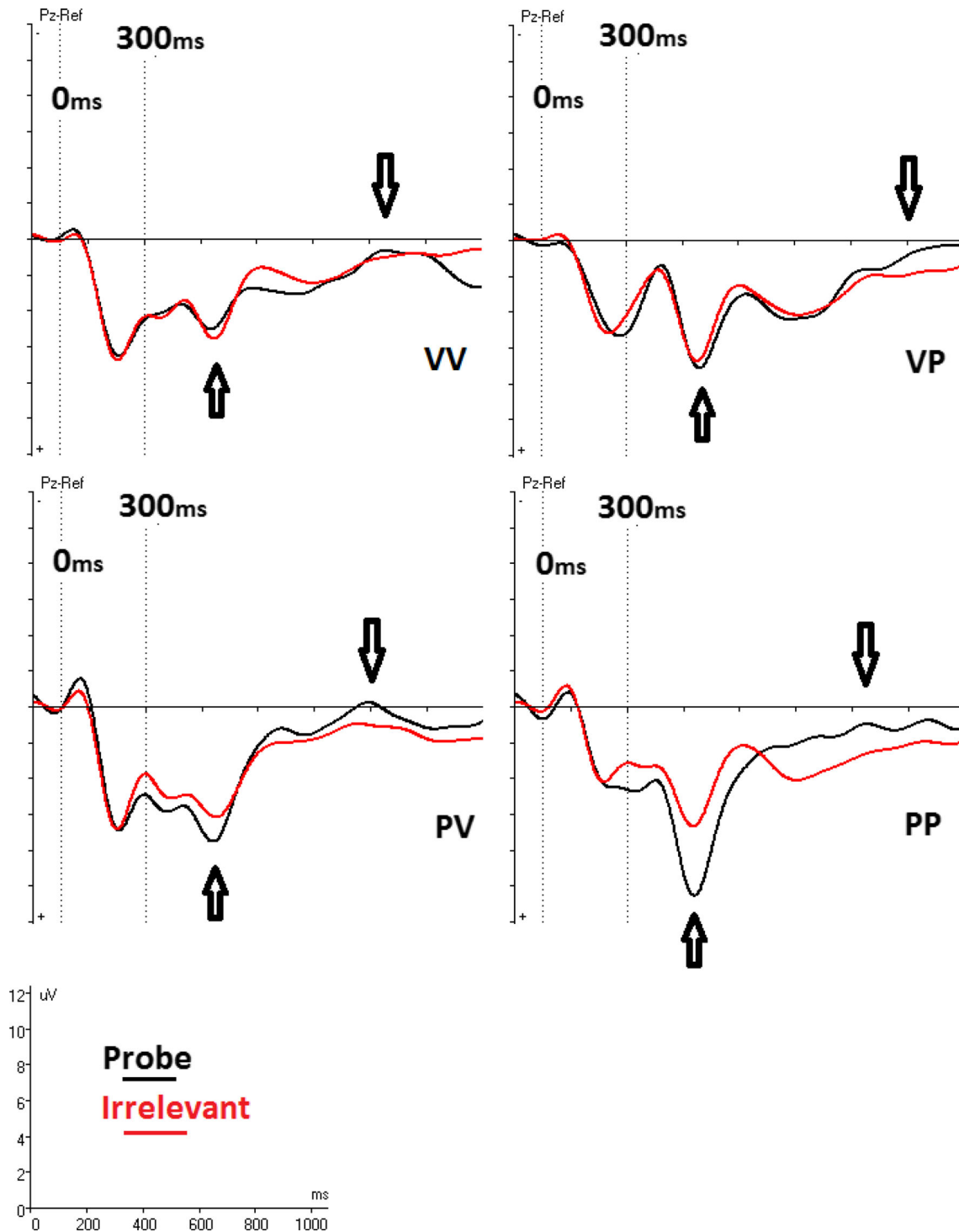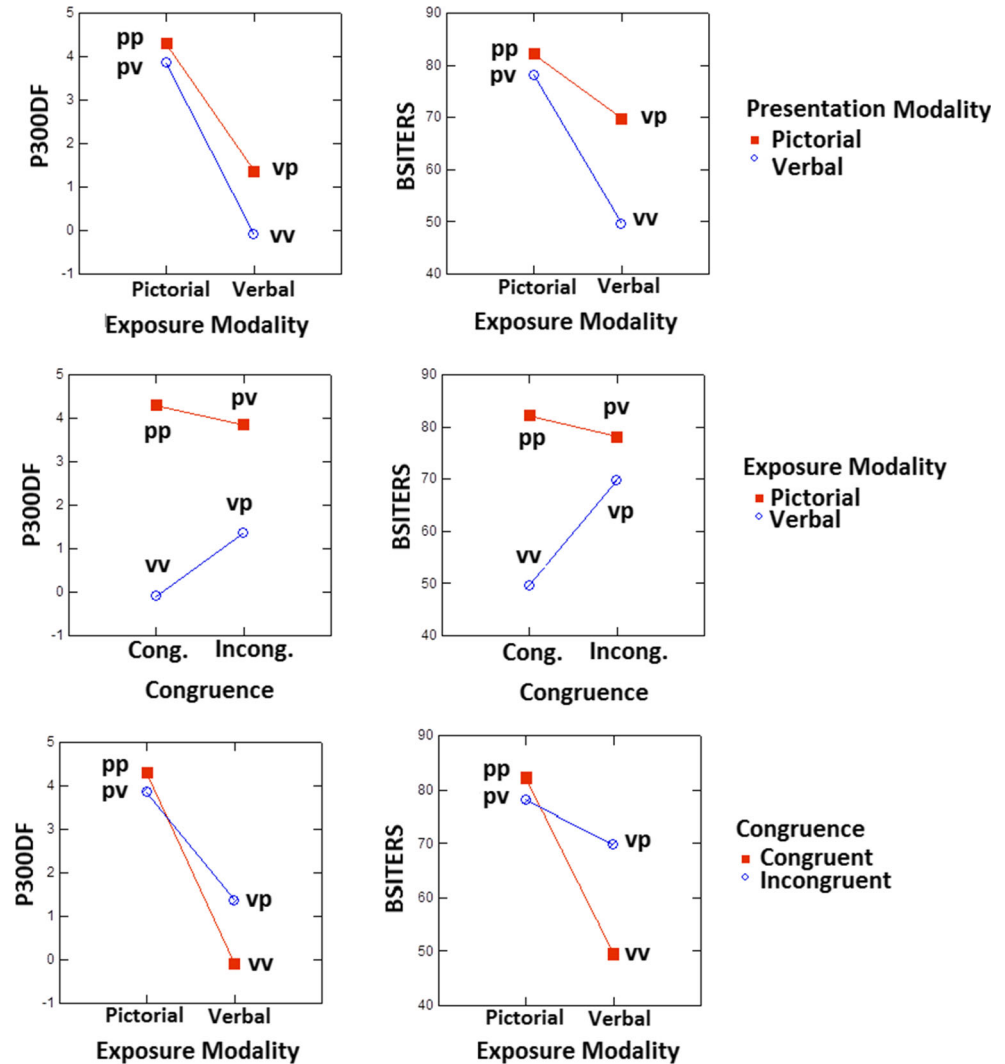
Fig. 3 Superimposed Pz probe (*black*) and irrelevant (*red*) grand average waveforms sorted by various exposure modality–presentation modality combinations: PP and VV are congruent pictorial and verbal modalities, PV and VP are incongruent exposure–presentation modalities, with exposure being the first and presentation the second letter. *Up arrows* show peaks of P300 components, and *down arrows* show peaks of negative components following P300 but within the 1300 ms look window; peak differences yield the p–p P300 values. *Vertical faint dotted lines* at 0 and 300 ms show the stimulus onset and offset (Color figure online)

impression of Fig. 3 that PP > PV > VP > VV. The bootstrapped estimates of Pz, probe–minus–irrelevant p–p P300 differences in their populations (BSMEAN) are not

shown as they closely follow P300DF, a single sample estimate of the same value. However they will be analyzed in the quantitative section below. It is also seen in the left

Fig. 4 *Line graphs* plotting computed Pz p–p P300 probe–minus–irrelevant amplitude differences (P300DF) and number of Pz p–p P > I differences out of 100 bootstrap iterations (BSITERS), as a function of (exposure modality, presentation modality, and exposure congruence). In *each of the three panels*, the same exposure–presentation combinations (pp, pv, vp, vv) are plotted in different ways

columns of Fig. 4 that p–p Pz differences of probe minus irrelevant (P300DF; as expected in these all knowledgeable groups) are >0 except perhaps, for the VV condition, and it is also seen in the right columns of Fig. 4 that the numbers of P > I differences in the bootstrap tests (BSITERS; as expected and as usual) are >50/100 except perhaps, for the VV condition. This result predicts the usual main effect of stimulus type, probe P300 > irrelevant P300 in knowledgeable subjects.

ERP Results, Quantitative

For each of the dependent variables used here (P300DF, BSMEAN, BSITERS), a mixed, 3-way ANOVA was performed with the between-subject factor of exposure modality (pictorial vs. verbal), the other, repeated measure, key variable of interest, presentation modality (pictorial vs. verbal), and the repeated measure, nuisance variable, day (1 vs. 2; see Table 1). The five significant and one not quite

significant ($p < .07$) results are tabulated in Table 2, excepting the nuisance variable, day, which had no significant main effects nor interactions with any other variables. Also, there were no significant 2-way or 3-way interactions; all $p > .2$, excepting the interaction of presentation modality and exposure modality for BSITERS (upper right, Fig. 4), which was not significant at $F(1,9) = 3.3$, $p = .102$, $\eta_p^2 = .27$.

In order to verify the expected main effect of probe > irrelevant (in these all knowledgeable subjects) across all exposure and presentation modalities, we performed another ANOVA, on separated probe and irrelevant P300 data (P300DF only). This was a 4-way ANOVA with exposure again as a factor; and day, presentation modality, and stimulus type (probe vs. irrelevant) as repeated measures. The expected main effect of stimulus type did obtain, $F(1,9) = 7.49$, $p < .03$, $\eta_p^2 = .45$. The only other significant effect was the interaction of stimulus type with presentation modality; $F(1,9) = 5.24$, $p < .04$, $\eta_p^2 = .40$.

Table 2 ANOVA data (*Modality* is shortened here to *Mode*)

Measure	Main effect	ANOVA data	Power[a]
P300DF	Exposure mode	$F(1,9) = 3.98$, $p = .07$, $\eta_p^2 = .31$, $\eta^2 = .182$	.77
	Presentation mode	$F(1,9) = 5.49$, $p < .05$, $\eta_p^2 = .38$, $\eta^2 = .014$	.81
BSMEAN	Exposure mode	$F(1,9) = 4.66$, $p = .05$, $\eta_p^2 = .34$, $\eta^2 = .194$	.74
	Presentation mode	$F(1,9) = 5.44$, $p < .05$, $\eta_p^2 = .38$, $\eta^2 = .018$	.81
BSITERS	Exposure mode	$F(1,9) = 4.45$, $p < .05$, $\eta_p^2 = .38$, $\eta^2 = .178$	.81
	Presentation mode	$F(1,9) = 5.44$, $p < .05$, $\eta_p^2 = .38$, $\eta^2 = .018$	.81

[a] "Post hoc achieved power" as computed from *G*Power 3.1.9.2*

Probe p–p Pz P300 was always greater than irrelevant P300, but the difference was greater in the pictorial condition (9.32 vs. 6.34 μV), than in the verbal condition (10.41 vs. 8.36 μV). The interaction of exposure modality with stimulus type was close to significant, $F(1,9) = 3.93$, $p < .08$, $\eta_p^2 = .23$. Again, the probe minus irrelevant condition was greater for the pictorial than the verbal exposure. These main effects and interactions for stimulus type are all suggested by Figs. 2 and 4.

From another point of view, exploring effects of exposure–presentation *congruence* on P300DF in a 3-way ANOVA, involving the factor of exposure modality, and the repeated measures variables, day and congruence, yields a congruence by exposure interaction identical to the main effect of presentation modality in Table 2: $F(1,9) = 5.49$, $p < .05$, $\eta_p^2 = .38$, as these analyses are mathematically redundant. The congruence by exposure interaction is evident in the lower two panels of Fig. 4, where there is a crossover of the two lines in both panels. Likewise for congruence by exposure interactions on BSMEAN and BSITERS. Graphical redundancy occurs also as the top and bottom panels of Fig. 4 involve the same data points connected differently in comparing the top and bottom panel pairs.

Discussion

It is evident that both exposure and presentation modality affect sensitivity in the P300-CIT version called the CTP. It would appear superficially that presentation has a greater effect than exposure, as the overall large effect size values (for η_p^2, a *large* effect $>.14$, according to Cohen 1988, as described also in Richardson 2011) are consistently larger (by .04–.08) for presentation than for exposure. However, as noted above, Richardson (2011) also demonstrated that one cannot compare η_p^2 values for independent variables in a mixed, higher order factorial design, recommending that for such comparative purpose the use of η^2. With these latter effect size values, Table 2 shows that the exposure modality effect size has a consistently much larger value than that of the presentation modality effect size. This is

quite consistent with the well-known behavioral "Pictorial Superiority" effect studied extensively by Stenberg (2006) and Stenberg et al. (1995), and known about for 100 years (Kirkpatrick 1894). This effect was also recently extended to the ERP domain, for example, by Curran and Doyle (2011) who noted that their "old-new effect" seen in late parietal potentials is larger for pictorial than verbal exposure. As noted above, the parietal "old-new" effect seen in late positive potentials is not necessarily the same thing as the P300 oddball effect, however there are similarities—indeed some (e.g., Friederici et al. 2001) argue that the parietal "old-new" effect is indeed a P300 after all. Indeed, the fact that the parietal "old-new" effect parallels what we here report as a significant effect of exposure modality favoring pictorial exposure on P300 probe–irrelevant difference amplitudes, provides support for the present findings.

Stenberg et al. (1995) also report an exposure-test *congruence* effect in their first two experiments, as do many others as summarized in Stenberg (2006). However the same authors also note that certain manipulations yield a pictorial *presentation* modality effect which overcomes the exposure–presentation *congruence* effect (Stenberg et al. 1995 in their experiments 3 and 4): For example in their fourth study, the exposure modality *combined* pictorial and verbal elements, as pictures of study objects were accompanied by their names below the pictures, and the result was that the pictorial test (*presentation)* modality effect favoring pictures overshadowed the congruence effect they saw in their first two studies. We suggest that our similar result of pictorial test (presentation) superiority is based on our combining two blocks of exposure (one verbal, the other pictorial) in analyzing the effects of presentation. Madigan (1983) also reported pictorial testing superiority rather than a congruence effect.

The finding here of a presentation modality effect favoring pictorial over verbal presentation suggests in any case that CIT investigators utilize pictorial presentation whenever possible in CITs. Indeed, although common sense (and some literature, e.g., Graf et al. 1985; Tversky 1969; Heckler and Childers 1992; Peeck 1974; Stenberg et al. 1995) might suggest that consistent exposure–

presentation modalities be employed in CIT testing, in the present data set, this turns out not to be a good idea, since pictorial presentation was always more effective regardless of exposure modality. This is suggested by the upper four panels of Fig. 4.

A matter needing comment here are the relatively low values of BSITERS seen here in the right columns of Fig. 4, which range from 49 (VV) to 82(PP). In previous P300 CIT studies, these values are rarely <75, and usually in a range from 80 to 95 (reviews in Rosenfeld 2011; Rosenfeld et al. 2013). We suspect that one major reason for the lower values seen here relates to the presently controlled, initial exposure to stimuli: In our previous P300 mock crime studies, probe stimuli from mock crimes were directly handled and seen by participants who concealed these items on their persons during all 40 min of the mock crime and subsequent testing; (Rosenfeld 2011; Rosenfeld et al. 2013). In contrast, because here we wanted to control the visual perceptual parameters such that verbal and pictorial stimulus field subtended the same retinal angles, subjects saw identically sized verbal as well as pictorial probe stimuli on a two dimensional computer screen for 30 s, followed by a 30 s mental imaging period, followed by a final display screen exposure of 30 s. It is quite conceivable that this later, un-natural probe exposure resulted in a more superficial depth of processing and rehearsal mechanism than what would be encountered in a real or mock crime, and thus was also likely to result in poorer memory encoding and correspondingly smaller values of its subsequent P300 sign of recognition. This is indeed a limitation of the current study regarding ecological validity, but it seemed the most efficient method to exert the control we wanted so as to be able to better answer the modality questions in an un-confounded manner, as our present focus was not ecological validity.

Another possible explanation for the lower bootstrap scores (BSITERS) in this study may be due to the 1:1 target/non-target ratio we used in part 2 of the present CTP trial. That is, the probability of trials containing targets was .5 in this present study. In other studies, however, in which we also used symmetric conditional probability of target following probe and irrelevant, but in which ratio of target to non-target was much lower, rather high accuracies (BSITER scores) were obtained (e.g., 100 % classification accuracy and 1.0 ROC area in Meixner and Rosenfeld (2011), who used a 1/9 target to non-target ratio, i.e., a target probability of .1.) A unique response-requiring target appearing on half of all trials, as in the present report, would tend to divide subjects' attention toward the target task and away from the probe–irrelevant discrimination task, thus reducing probe P300 amplitude and thus accuracy. (In fact, two present participants made us aware of the just noted effect by complaining about "having to watch for targets all the time".) Of course, the aim of this report was not to maximize accuracy but to compare exposure and presentation modality effects, so our use of an overall target probability of .5 was probably of no consequence.

What is the basis of the pictorial superiority effect regarding exposure? The present study bears on this question, although with its emphasis on applied questions, was not intended to deal in detail with complex memory mechanisms. (One clear complexity is that the processes underlying pictorial *exposure* superiority must involve some differences in comparison to the processes underlying pictorial *presentation* superiority.) Nevertheless, it has often been suggested that a fundamental variable for both exposure and presentation effects is stimulus complexity, perceptual and conceptual (Stenberg et al. 1995; Stenberg 2006). Pictures have more perceptual details in complex arrangements than do words, so that exposed pictorial material may be better remembered than exposed verbal material because there is more opportunity to match complex perceptual information to a template with complex pictures, than to match relatively barren verbal material to a verbal template. Color, for example, is a perceptual attribute possessed by our pictorial but not verbal stimuli, even suggesting a possibly confounding condition here: verbal stimuli were monochromatic, whereas pictorial stimuli were color photographs of objects. This confound probably does not account for the pictorial exposure superiority observed here, since Nelson et al. (1974) compared recognition for color photographs versus embellished line drawings versus simple line drawings, versus words, and found no differences among pictorial stimuli which were all better remembered than words. Nevertheless, there are attributes other than perceptual attributes such as color, which contribute to the complexity of pictures but not words. In any case, Stenberg et al. (1995, p. 437) did suggest that their findings were "consistent with better semantic access from pictures than from words" as an explanation of the pictorial superiority effect. Pictures were said to have a "privileged access…to a common semantic store." However, the reasons for this privileged access were not clear to Stenberg et al. (1995). A decade later Stenberg (2006) expanded the pictorial superiority effect to possibly include both conceptual and perceptual processing advantages for pictures, but noting that the former makes a greater contribution for pictorial memory than for verbal memory. It is suggested in conclusion that pictorial exposure superiority seen here with P300 is probably closely related to the effects seen in the behavioral memory literature and is probably mediated by the same recognition processes in view of the fact that P300 is a recognition index in the CIT.

In summary, the present results are mainly consistent with previous behavioral memory studies and their

extension to electrophysiological studies. They strongly urge that pictorial test presentation modality be used whenever possible in P300-based tests for concealed information.

Conflict of interest The authors have no potential conflicts of interest regarding this paper.

Human Rights and Informed Consent Human subjects were used. Informed Consent was obtained from all subjects and the project was approved by the Northwestern University IRB.

References

Allen, J. 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*(5), 504–522. doi:10.1111/j.1469-8986.1992.tb02024.x.

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A concealed information test with multimodal measurement. *International Journal of Psychophysiology, 75*(3), 258–267.

Ben-Shakhar, G., Frost, R., Gati, I., & Kresh, Y. (1996). Is an apple a fruit? Semantic relatedness as reflected by psychophysiological responsivity. *Psychophysiology, 33*, 671–679.

Ben-Shakhar, G., & Gati, I. (2003). The effects of serial position and frequency of presentation of common stimulus features on orienting response reinstatement. *Psychophysiology, 40*, 1–7.

Childers, T. L., Houston, M. J., & Heckler, S. E. (1985). Measurement of individual differences in visual versus verbal information processing. *Journal of Consumer Research*, 125–134.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Curran, T., & Doyle, J. (2011). Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience, 23*(5), 1247–1262.

Cutmore, T. R., Djakovic, T., Kebbell, M. R., & Shum, D. H. (2009). An object cue is more effective than a word in ERP-based detection of deception. *International Journal of Psychophysiology, 71*(3), 185–192.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology, 41*(5), 665–678.

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

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology, 28*(5), 531–547. doi:10.1111/j.1469-8986.1991.tb01990.x.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*, 305–323.

Gehring, R. E., Toglia, M. P., & Kimble, G. A. (1976). Recognition memory for words and pictures at short and long retention intervals. *Memory and Cognition, 4*(3), 256–260.

Graf, P., Shimamura, A. P., & Squire, L. R. (1985). Priming across modalities and priming across category levels: Extending the domain of preserved function in amnesia. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*(2), 386.

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

Heckler, S. E., & Childers, T. L. (1992). The role of expectancy and relevancy in memory for verbal and visual information: What is incongruency? *Journal of Consumer Research*, 475–492.

Herron, J. E., & Rugg, M. D. (2003). Retrieval orientation and the control of recollection. *Journal of Cognitive Neuroscience, 15*(6), 843–854.

Introduction to SAS. UCLA: Statistical consulting group. http://www.ats.ucla.edu/stat/sas/notes2/. Accessed November 24, 2007. Also, http://www.ats.ucla.edu/stat/mult_pkg/faq/general/effect_size_power/effect_size_power.htm

Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology, 30*(1), 90–97.

Keil, A., Debener, S., Gratton, G., et al. (2014). Committee report: Publication Guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology, 51*(1), 1–21.

Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychological Review, 1*(6), 602.

Labkovsky, E., & Rosenfeld, J. P. (2014). A novel dual probe complex trial protocol for detection of concealed information. *Psychophysiology,*. doi:10.1111/psyp.12258.

Lui, M., & Rosenfeld, J. P. (2008). Detection of deception about multiple, concealed, mock crime items, based on a spatial–temporal analysis of ERP amplitude and scalp distribution. *Psychophysiology, 45*(5), 721–730.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43*(6), 385.

Madigan, S. (1983). Picture memory. In J. C. Youille (Ed.), *Imagery, memory, and cognition*. Hillsdale, NJ: Erlbaum.

Meijer, E. H., Smulders, F. T., Johnston, J. E., & Merckelbach, H. L. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology, 44*(5), 814–822.

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300 based concealed information test. *Psychophysiology, 48*(2), 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.

Nelson, T. O., Metzler, J., & Reed, D. A. (1974). Role of details in the long-term recognition of pictures and verbal descriptions. *Journal of Experimental Psychology, 102*(1), 184–186.

Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: Evidence for pictorial superiority. *Journal of Gerontology, 38*(5), 582–588.

Peeck, J. (1974). Retention of pictorial and verbal content of a text with illustrations. *Journal of Educational Psychology, 66*(6), 880.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 63–89). New York, NY: Cambridge University Press.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. H. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology, 28*(3), 319–335. doi:10.1111/j.1469-8986.1991.tb02202.x.

Rosenfeld, J. P., Ben-Shakhar, G., & Ganis, G. (2012). Detection of concealed stored memories with psychophysiological and neuroimaging methods. In L. Nadel & W. Sinnott-Armstrong (Eds.),

Memory and law (pp. 263–303). Oxford: Oxford University Press.

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

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

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

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*(6), 906–919. doi:10.1111/j.1469-8986.2008.00708.x.

Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological Psychology, 74*(3), 396–404. doi:10.1016/j.biopsycho.2006.10.002.

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology, 41*(2), 205–219. doi:10.1111/j.1469-8986.2004.00158.x.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Ccognitive Sciences, 11*(6), 251–257.

Schacter, D. L., & Graf, P. (1989). Modality specificity of implicit memory for new associations. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*(1), 3.

Schupp, H., Cuthbert, B., Bradley, M., Cacioppo, J., Ito, T., & Lang, P. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology, 37*, 257–261.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23*(6), 695–703.

Seymour, T. L., & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology, 22*(4), 475–490.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156–163.

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.

Stenberg, G., Radeborg, K., & Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. *Memory and Cognition, 23*(4), 425–441.

Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology, 18*(6), 813–847.

Tversky, B. (1969). Pictorial and verbal encoding in a short-term memory task. *Perception and Psychophysics, 6*(4), 225–233.

Weinberg, A., & Hajcak, G. (2011). The late positive potential predicts subsequent interference with target processing. *Journal of Cognitive Neuroscience, 23*(10), 2994–3007.

Wells, J. E. (1972). Encoding and memory for verbal and pictorial stimuli. *The Quarterly Journal of Experimental Psychology, 24*(3), 242–252.