Appl Psychophysiol Biofeedback (2009) 34:177–187
DOI 10.1007/s10484-009-9092-3

Time and Encoding Effects in the Concealed Knowledge Test

Travis L. Seymour · Becky R. Fraynt

Published online: 18 June 2009

Abstract Although the traditional "lie detector" test is used frequently in forensic contexts, it has (like most test of deception) some limitations. The concealed knowledge test (CKT) focuses on participants' recognition of privileged knowledge rather than lying per-se and has been studied extensively using a variety of measures. A "guilty" suspect's interaction with and memory of crimescene items may vary. Furthermore, memory for crimescene items may diminish over time. The interaction of encoding quality and test delay on CKT efficiency has been previously implied, but not yet demonstrated. We used a response-time based CKT to detect concealed knowledge from shallow and deep study procedures after 10-min, 24-h, and 1-week delays. Results show that more elaborately encoded information afforded higher detection accuracy than poorly encoded items. Although classification accuracy following deep study was unaffected by delay, detection of poorly elaborated information was initially high, but compromised after 1 week. Thus, choosing optimal test items requires considering both test delay and initial encoding level.

Keywords Guilty knowledge test · Concealed knowledge test · Response time measure · Detection of deception · Lie detection · Applied psychology · Law policy · Levels of processing

T. L. Seymour (✉)
Psychology Department, 357 Social Sciences 2, University of California, Santa Cruz, Santa Cruz, CA 95064, USA
e-mail: nogard@ucsc.edu

B. R. Fraynt
University of California, Los Angeles, CA, USA

Introduction

Despite the popularity of traditional "lie detector" tests and numerous reports of their success (for a review, see Honts et al. 2005), several limitations have been identified (e.g., Lykken 1998; National Research Council 2003; Raskin 1989). One promising alternative is to measure *concealed knowledge*. Instead of relying on suspects feeling aroused or anxious when deceptively answering crime-related questions (e.g., "Did you shoot the drugstore guard on June 23rd?"), the concealed knowledge test (CKT) (Lykken 1959) indexes an examinee's recognition of crime-relevant information. The typical CKT[1] paradigm presents a critical *probe* stimulus alone with several *irrelevant* items. For example, "The person who stole the statue would recall its appearance. Was it made of (a) gold, (b) silver, (c) wood, (d) glass, or (e) plastic?" Participants are asked to respond "No" after each answer choice is presented. During this process, one or more physiological measures are recorded, and differential responsivity to probe choices compared to irrelevant alternatives indicates knowledge of the crime. The CKT has been successfully coupled with a variety of physiological measures such as heart-rate, electrodermal response (EDR) (for review see Ben-Shakhar and Elaad 2003), brain electrical activity (e.g., Rosenfeld et al. 1988), and pupil dilation (e.g., Lubow and Fein 1996). More recently, it has also been successful with behavioral measures such as response time and accuracy (e.g., Allen et al. 1992; Seymour and Kerlin 2008; Seymour et al. 2000).

Time and Encoding Effects on the CKT

Despite its success in the laboratory, surprisingly few studies have investigated the time course of the concealed knowledge effect. This is particularly important because in

applied contexts the delay between exposure to privileged information and the subsequent test may be quite variable, ranging from minutes to years. In previous studies, the effect of time on CKT efficiency can vary from one dependent measure to another. For example, one study used an EDR-based CKT and examined detection accuracy for tests immediately following a mock-crime as well as those delayed by 1 week (Carmel et al. 2003). These results revealed no effect of delay for either EDR magnitude or detection accuracy. Elaad (1997) also reported successful EDR-based tests after 1 week. Using an event-related brain potential (ERP) based CKT, Rosenfeld et al. (1991) examined tests that either immediately followed probe study or were separated by 7–14 days. They reported high accuracy on immediate tests similar to previous reports (e.g., Farwell and Donchin 1991; Rosenfeld et al. 1988), but not in delayed conditions. Accurate detection in delayed conditions required re-exposing participants to critical items immediately prior to test.

The success of both immediate and delayed tests may also be influenced by the level of attention allocated to each stimulus at initial exposure, and the resulting quality of memory for these items. Information present at the crime scene but never encoded will be unfamiliar and decrease the accuracy of a CKT using such information. Indeed, most CKT procedures explicitly correlate lack of response to test items with lack of prior exposure. For example, Carmel et al. (2003) reported that when participants were alerted to which mock-crime details were important, later EDR-based CKT accuracy was higher than when noticing probe items was left to chance. Regardless of whether they were made explicit, using probes that were more central to a crime (e.g., the weapon) led to higher detection accuracy than more peripheral items (e.g., veneer of a table). Unfortunately, it is not easy to determine whether the decrease in accuracy reported by Carmel and colleagues for un-cued crime items was due to participants' poor encoding of the items or their failure to encode them altogether. This is an important distinction because it may be possible to detect poorly encoded information, but not unencoded information.

A more straightforward approach to answering this is to have participants explicitly study probe items and then manipulate their degree of encoding. Thus, CKT accuracy could be evaluated for well-encoded or poorly encoded probes while minimizing the likelihood that they are completely unencoded. For example, a recent study compared ERP-based CKTs in which the only probe item was either the participant's name (highly elaborated in participants' memory) or the experimenter's name (newly learned) and showed greater classification accuracy for the more elaborated stimulus (Rosenfeld et al. 2006). Although this paradigm successfully varies levels of probe encoding, the use of a single probe item with such a special status as one's own name makes it an effective, but somewhat extreme demonstration. A variation on the Rosenfeld et al. (2006) paradigm would be to explicitly expose participants to either a set of probe items that they encode richly or ones that they encode poorly, but avoiding the special status of one's name. Previous studies have demonstrated that the amount of attention paid to a particular crime item is positively correlated with both the level of physiological response to that item during a lie test, as well as to the item's later explicit recall by the participant. In these studies, level of attention was not manipulated, but inferred by the level of physiological response to each stimulus during study (e.g., Waid et al. 1978, 1981a). This correlation between attention and detectability appears to predict that using peripheral items in the CKT will lead to poorer detection efficiency than more central items. However, though Waid et al. (1981a, b) were able to find this relationship using the traditional "lie-detector" test, they did not find that attention and EDR were strongly correlated in the CKT.

Related non-applied work has also examined the relationship between the manner in which material is processed and the likelihood that it will be later recalled. Craik and Lockhart's (1972) Levels of Processing (LOP) approach typically contrasts shallow (e.g., focusing only on the superficial visual characteristics of a word) and deep processing (e.g., focusing on the category membership or meaning of a word) of stimuli. Although various encoding types may lead to some long-term storage, the probability of later recall was higher for more deeply processed stimuli. As Baddeley (1999) points out, deep processing may also refer to the richness or breadth of encoding, so that focusing on multiple aspects of an item (e.g., visual, auditory, and semantic) may enhance its later recall compared to focusing on only one aspect. Thus, crime-scene information leading to more elaborated memories (viz., items used to commit the crime, or highly salient aspects such as the victim's face) should be more successful probe items than less salient details (e.g., flowers growing at the crime scene; Nakayama 2002).

Do Time and Encoding Interact in the CKT?

Although by some accounts six or more specific and central probe items may be ideal for an effective CKT (e.g., Lykken 1998), in forensic investigations key case details can often be leaked or otherwise compromised (Podlesny 2003). In such cases it may be tempting to consider CKT probe items based on peripheral crimescene information. Unfortunately, the accuracy of the CKT using peripheral information, especially after long delays, is unclear. Related issues have been explored in studies on eyewitness

memory and typically show that central crime details are better remembered than peripheral ones when tested after a delay. Central details have also proven more resistant to alteration by later misleading information (e.g., Christianson and Loftus 1991; Loftus 1979). Although, it seems advisable to limit CKT probes to the most central crime details, it is still possible that peripheral items will be sufficiently recollected for detection. Clearly, a concealed knowledge effect cannot be observed for information absent from memory (Carmel et al. 2003), but poorly elaborated memories may still be detected by the CKT.

Because later memory for crime items may be influenced by an interaction between encoding effects at the crimescene and the time between crime and test, it may be difficult to determine the appropriateness of a particular probe set by considering either quality of encoding or test delay alone. However, to our knowledge, no previous study has simultaneously examined the influence of these variables on CKT efficiency. In the present study we examined this interaction using a variation of the CKT procedure that uses response times (RT) and accuracy as measures (Seymour and Kerlin 2008; Seymour et al. 2000). Although Seymour and colleagues have shown that an RT-based paradigm can accurately detect concealed knowledge with immediate tests, this paradigm has yet to be evaluated in delayed tests. Benefits of this paradigm include its ease and low cost of measurement compared to some psychphysiological measures, as well as a more straightforward analysis procedure.

Methods

We used an RT-based CKT under two conditions: (a) probe items were elaborated during study using multiple stimulus and response modalities; (b) probe items were minimally elaborated during study. In addition, we varied the time delay between probe study and the later test.

Participants

Participants were 109 undergraduate students (60 female) recruited by flyer from the University of California Santa Cruz community. Participants were promised $15 for their participation, and were later offered an additional $10 incentive to "beat the test."

Materials

Stimuli consisted of 72 two-word phrases previously used in this and similar paradigms (Farwell and Donchin 1991; Rosenfeld et al. 2004; Seymour and Kerlin 2008; Seymour et al. 2000). For each participant, phrases were randomly arranged into two sets of 36 phrases that contained six subgroups; names (e.g., "Phil Jenks"), street names (e.g., "Perch Street"), file descriptions (e.g., "Rain File"), articles of clothing (e.g., "Blue Coat"), and operation names (e.g., "Op Horse"). Within each stimulus set, six items (one from each category) were randomly designated as *probe* items, six items (one from each category) were randomly designated as *target* items, and 24 (four from each category) were randomly designated as *irrelevant* items.

Procedure

The procedure (depicted in Fig. 1) replicated previously reported RT-CKT paradigms and consisted of a probe study task, a delay period, and two concealed knowledge tests. Each test consisted of a target study task, and a phrase classification task. Participants were randomly assigned to either a *shallow* or *deep* probe-study condition, and to one of three delay conditions: 10 min, 24 h, or 1 week.

Probe Study Tasks

Participants in both study conditions first completed a cued-recall task in which they were asked to memorize the probe phrases. This task involved studying the six probes in a randomly ordered list (e.g., "First Phrase: Blue Coat") and attempting to commit them to memory. Participants were subsequently instructed to recall the list in order in response to positional cues (e.g., "First Phrase:") and were given accuracy feedback on their performance. This cued-recall task (study and test) was repeated three times.

Following cued-recall, the procedure diverges for participants in the shallow and deep conditions: those in the shallow condition completed a single news paraphrase task, whereas those in the deep study condition completed a series of probe association tasks.

The news paraphrase task required participants to read a mock newspaper story (140 words) that included the six probe phrases in the description of a fictitious campus theft and were then asked to paraphrase it in writing. Participants were informed that they would need to paraphrase the story and thus spent 3 min on average studying the text.

Instead of the newspaper task, participants in the deep condition completed a set of four new tasks (picture matching, word jumble, hand writing, and word shouting) designed to increase exposure duration, require increased attention, and lead to rich and multimodal memory representations (c.f., Waid et al. 1981b). Each of the four tasks involved random presentation of all six probes in three separate blocks for a total of 12 task blocks. These task blocks were randomly mixed together so that no two successive blocks involved the same task (e.g., pm, wj, ws,

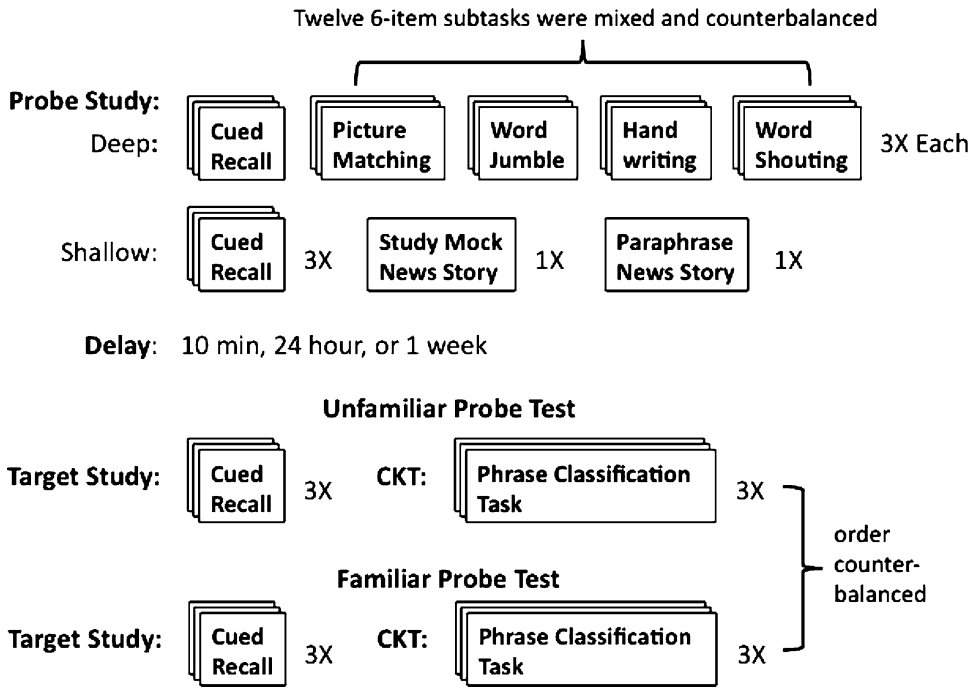

Fig. 1 Overview of experimental design

pm, hw, wj, etc.). After each response of each task, participants were given accuracy feedback, except for the shout task during which response time feedback was displayed instead. Overall, this task is similar to the *overlearning* technique previously reported by Waid and colleagues, who also asked participants to write each word backwards, provide the list in alphabetical order, and to provide free associates of each word.

During the word jumble task each probe phrase was presented with its letters strategically rearranged to obscure the source phrase (e.g., "Lion Street" became "ettlesin-or"). Participants were asked to type the phrase indicated by the jumbled string. For each block of the jumble task we presented a different jumble of each phrase with each letter arranged for maximum incongruity.

The picture matching task involved presentation of an image that referred to one of the probe phrases (e.g., a picture of a blue coat for "blue coat"). Participants were asked to verbally identify which phrase corresponded to the picture. Each block of the picture matching task used a different type of image; full color photorealistic images, sparsely colored line drawings, and grayscale illustrations (See Fig. 2).

On each trial of the word shouting task a probe phrase was presented and participants were asked to shout each phrase three times quickly. Emphasis was placed on responding quickly while still fully enunciating each word. Unlike the other probe association tasks, all three blocks of the word shouting task involved the same stimuli and responses.

Finally, during the handwriting task we presented each probe phrase and asked participants to write this phrase on paper. Each written response was to completely fill a 6-inch by 2-inch box pre-drawn on the paper. This constraint increased participants' engagement with each response and ensured legibility. Each handwriting block presented the same stimuli, but required a different type of writing; cursive, lowercase print, or uppercase print.

Test Delay

Following the set of probe study tasks, participants in the 10-min delay condition completed a distractor task designed to occupy working memory and prevent rehearsal of the probe items. The task consisted of 11 challenging mathematical word problems taken from Patalano and Seifert (1994). Participants in the 24 h and 1-week delay conditions were instead told that the session was over and asked to return to the lab after exactly 24 h or 7 days, respectively.

Target Study

After the delay, participants learned a new set of six target phrases by using the cued-recall procedure described above for the probe-study task. However, none of the additional study tasks used during probe study (i.e., neither newspaper nor overlearning tasks) were used during target study. Thus, regardless of whether participants were in the shallow or deep probe-study condition, target study always

Fig. 2 Examples of how images differed for each picture matching task block in the Deep Study condition (note: *colorful images* are shown here in *grayscale*). Presentation of each probe concept (e.g., "snow file" or "sub plans") was depicted using a full color photograph (*top row*), a sparse color sketch (*middle panel*), and a grayscale illustration (*bottom panel*)

consisted of just the cued-recall procedure. This ensures that target items were always less elaborated than probes and this arrangement mimics the relationship between targets and probes one might expect in forensic use of this paradigm. To the knowledgeable participant in such contexts, probe items should be more elaborated and consolidated in memory compared to target items learned just prior to test (Elaad 1997). This would be true even if the examiner took care to match probe and target items on appearance and conceptual makeup (e.g., "White Shirt" and "Brown Pants").

Classification Task

After the target-study task, participants performed a series forced-choice binary classifications consisting of 6 targets, 6 probes, and 24 new irrelevant phrases in randomized order. Participants were asked to indicate their familiarity

with each item by pressing buttons labeled "Yes" and "No" for familiar and unfamiliar stimuli, respectively. They were asked to respond truthfully to familiar target items ("Yes") and new irrelevant phrases ("No"), but they were asked to respond deceptively to familiar probes items ("No"). Before each stimulus was displayed, the word "Ready" was displayed for 500 ms, followed by a fixation cross for 500 ms. The stimulus remained on the screen until a response was made, and both speed and accuracy were equally stressed. If responses were slower than 1,500 ms, the message "Too Slow" was displayed for 1 s before the next stimulus was displayed; otherwise, no feedback was given within blocks. The same test list was re-randomized and repeated three times for a total of 108 trials. After each test block (i.e., a sequence of 36 trials), participants were given feedback on accuracy as well as the number of "Too Slow" errors. On each trial, RT and accuracy were recorded using "E-Prime" stimulus presentation software (Schneider et al. 2002).

The combination of the target-study and phrase-classification task was completed twice for each participant in two separate tests, neither of which shared any of the same phrases. During the *unfamiliar-probe* test, probe phrases were novel and participants had no means of distinguishing probe and irrelevant items. Data from this test is equivalent to testing a participant who has no knowledge of crime details and served as a basis to estimate the test's false positive rate, which is typically between 0 and 3% (e.g., Seymour and Kerlin 2008; Seymour et al. 2000). Alternatively, probes in the *familiar-probe* test were taken from the earlier probe-study task and expected to result in slower and less accurate responses on probe than irrelevant trials.

Prior to completing these tests, we informed participants that the procedure was a new type of "lie-detector test" designed to measure their familiarity with probe items they previously learned. They were encouraged to try and conceal their knowledge of any familiar-probe phrases they encounter by responding just as quickly and accurately to probes as irrelevants. Although recruited with the understanding that they would receive $15 for the study, participants were now shown and offered an additional $10 bill for successfully concealing their knowledge of probe phrases. This bonus was contingent on achieving a typical unfamiliar-probe pattern of results: at least 85% accuracy on all trial types (target, probe, and irrelevant), and statistically identical probe and irrelevant RT distributions (assessed with a *t*-test). The 85% accuracy constraint rules out the strategy of ignoring the target stimuli and quickly pressing "No" on each trial. This is indeed the benefit of having target trials and distinguishes the present CKT paradigm from some others that only include probes and irrelevant items (for a review of several CKT variations, see Ben-Shakhar and Elaad 2003). Test order (familiar-probe vs. unfamiliar-probe) was counterbalanced and, participants learned a different set of target phrases for each test.

Results and Discussion

Three participants were not able to achieve at least 67% correct on the target or probe-study tasks even after three study iterations. Because we could not establish that these participants had memorized the test items, they were excluded from the study and subsequent analyses.

Familiar-Probe Test

Response Time

The effect of concealed knowledge on RT was calculated for each participant by subtracting mean irrelevant RT from mean probe RT (both "No" responses). The mean effect on RT in this test is shown in Fig. 3 (top graph) as a function of test delay and probe study condition, and was entered into a 2 (condition: deep, shallow) × 3 (delay: 10 min, 1 day, 1 week) ANOVA. Overall, the RT effect in the deep condition ($M = 115.55$; SD = 49.17) was greater than in the shallow condition ($M = 87.87$; SD = 39.25), $F(1, 100) = 10.23$, $p < 0.01$, $\eta_p^2 = 0.09$. A main effect of delay was also revealed, $F(2, 100) = 4.62$, $p < 0.02$, $\eta_p^2 = 0.08$, presumably driven by the decreasing effect in the shallow condition over time. Though the magnitude of the RT effect for deep study, but not the shallow-study condition, remained constant over time, the condition × delay interaction was not statistically significant, $F(2, 100) = 1.79$, $p = 0.17$, $\eta_p^2 = 0.03$. This lack of interaction was presumably influenced by the 24-h delay in the shallow condition.

Although the concealed knowledge effect differed between the 10 min and 24 h delays, $t(29) = 2.93$, $p < 0.01$, and the 10 min and 1 week delays, $t(37) = 4.44$, $p < 0.001$, the 24 h vs. 1 week delay difference was not, $t(34) < 1$. Thus, we performed an additional 2 condition × 2 delay ANOVA (omitting the 24 h delay group). This analysis resulted in main effects of Condition, $F(1, 73) = 9.91$, $p < 0.01$, $\eta_p^2 = 0.08$, and Delay, $F(2, 73) = 6.22$, $p < 0.02$, $\eta_p^2 = 12$, similar to the previous analysis. However, the interaction was now marginally significant, $F(2, 73) = 3.78$, $p = 0.06$, $\eta_p^2 = 0.05$.

Accuracy

The effect on accuracy was calculated by subtracting probe from irrelevant accuracy for each participant (see Table 1). An ANOVA was performed on these data similar to the RT analysis but revealed no main effects or interactions. In general, accuracy was near ceiling regardless of the quality

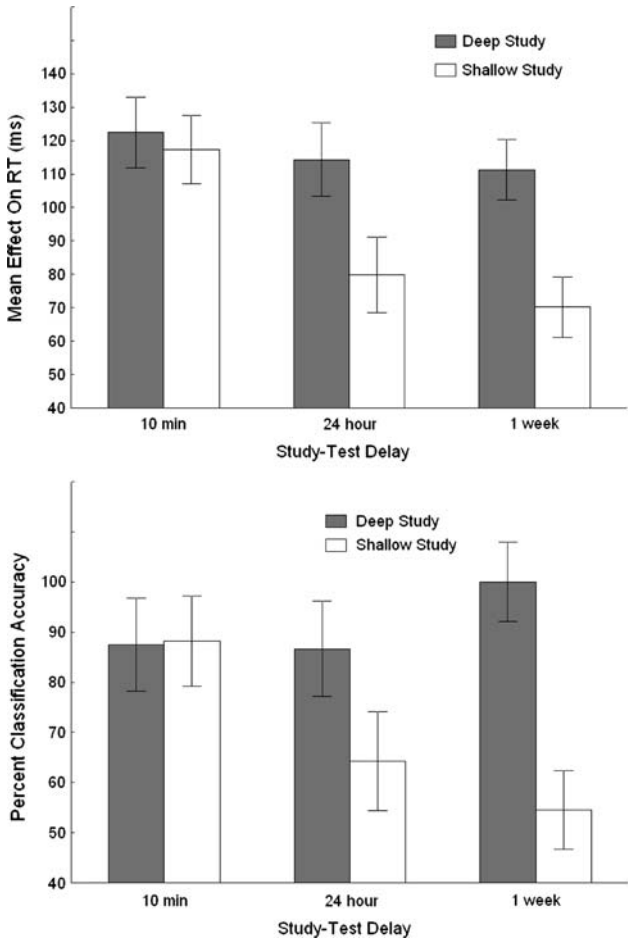

Fig. 3 *Top graph*: Mean concealed knowledge effect on RT in milliseconds (probe-irrelevant) for the familiar-probe test. *Positive values* indicate probe RTs were slower than irrelevant RTs. *Data* are shown as a function of study-test delay and probe study condition. *Error bars* represent ±1 SEM. *Bottom graph*: Proportion of correct classifications in the familiar-probe test (i.e., hit rate). *Classifications* are shown as a function of study-test delay and probe study condition. *Note*: three false positives (3.3% of participants) are not represented

of study or delay between study and test. Although this would be problematic in a standard recognition study where accuracy may serve as an indication of how appropriate a task is for testing participants' memory, here the result serves as a manipulation check on the financial incentive offered to participants. It also suggested that knowledge detection accuracy would be driven primarily by the RT measure.

Classification Analysis

An individual participant classification was used to differentiate participants' familiar-probe test from their unfamiliar-probe test. Because each participant completed familiar and unfamiliar-probe tests, a hit rate and false positive rate could be calculated for each participant. A three-part algorithm compared probe and irrelevant distributions on shape (Kolmogorov–Smirnov Test), variance (F-test for variances), and number of errors (Fisher's exact test). If either one of these tests was statistically significant compared to a Bonferonni corrected alpha level (i.e., 0.05 divided by 3 tests = 0.016), the probe and irrelevant distributions were assumed to be distinct, and thus the participant was deemed "familiar" with probe items during that test. If none of these statistical tests reached significance, it was concluded that the data were from a familiar-probe test (a hit for familiar-probe data, a false alarm for unfamiliar-probe data). Otherwise, it was assumed that the responses were from the unfamiliar-probe test (a miss for familiar-probe data, and a correct rejection for unfamiliar-probe data). This analysis has been used on similar data in previous studies and has yielded high hit rates and low false alarm rates (Seymour and Kerlin 2008; Seymour et al. 2000).

Table 1 Mean concealed knowledge effect on accuracy in percent correct (irrelevant-probe) for the familiar-probe test

Probe Study	Delay		
	10 min	24 h	1 week
Deep	3.3% (1.0%)	7.9% (1.2%)	3.3% (0.8%)
Shallow	4.3% (2.0%)	3.8% (1.0%)	3.2% (1.3%)

Positive values indicate that probe responses were less accurate than irrelevant responses. Data are shown as a function of delay and probe study condition

Values in parentheses are 1 ± SEM

Figure 3 (bottom graph) shows classification accuracy by Condition and Delay in the familiar-probe test (i.e., mean hit rate). The resulting detection accuracy was analyzed using a 2 (study type) × 3 (delay) ANOVA. Although there was no main effect of delay, $F(2, 100) = 1.06$, $p = 0.35$, $\eta_p^2 = 0.02$, the decreasing detection accuracy for the shallow condition was sufficient to yield a main effect of condition, $F(1, 100) = 9.23$, $p < 0.01$, $\eta_p^2 = 0.08$. Furthermore, the disparate detection patterns over time as a function of study type was supported by a significant interaction, $F(2, 100) = 3.63$, $p < 0.05$, $\eta_p^2 = 0.07$. This analysis was conducted on distributions of binary values (either a one or zero for each participant in each cell representing correct or incorrect classification) which typically violate the normality assumption of ANOVA. Thus, we also conducted a contrast analysis on the number of correct detections per cell. With this analysis, we are able to test the specific interaction evident in Fig. 3 (bottom graph); that deep study leads to a stable number of correct classifications over time, whereas shallow study leads to decreasing number of correct classifications (for a review of contrast analysis on frequency data, see Furr and Rosenthal 2003). A contrast analysis on this pattern was significant, $F(1, 100) = 18.20$, $p < 0.001$,

$\eta_p^2 = 0.15$, whereas a test of the prediction that study type and delay do not interact was not confirmed, $F(1, 100) = 1.52$, $p = 0.22$, $\eta_p^2 = 0.01$.

Unfamiliar-Probe Test

Mean concealed knowledge effect on RT (probe-irrelevant) and accuracy (irrelevant-probe) in the unfamiliar-probe test is shown in Table 2 as a function of delay and condition. A 2 (condition) × 3 (delay) ANOVA was performed on the RT data, revealing no main effects or interactions. This is the expected pattern during the unfamiliar-probe test because participants are unable to distinguish between probe and irrelevant stimuli. A similar analysis was performed on the accuracy data from the unfamiliar-probe test (see Table 2). Though the accuracy effect in this test ranged from 0% (24 h and 1 week) to only 1.3% (10 min), a statistically significant main effect of delay was revealed, $F(2, 100) = 4.91$, $p < 0.01$, $\eta_p^2 = 0.09$. Although it appears that the accuracy effect at the 24-h time delay was 5% greater for deep than shallow conditions, a Bonferroni corrected post-hoc *t*-test revealed no significant differences between cells. No other main effects or interactions were found for the effect on accuracy. The detection algorithm described above for the familiar-probe test was used to calculate the false positive rate during the unfamiliar-probe test. Because the probes in this test were unfamiliar, the test should have ideally determined all participants to be "unfamiliar." The test yielded an overall correct rejection rate of 96.7%, producing a false positive for three participants (two in the 10-min deep condition, and one in the 1-week shallow study condition). Thus, due to an insufficient number of false positives, the effect of study type and delay could not be analyzed for the unfamiliar-probe test.

Table 2 Mean concealed knowledge effect on RT in milliseconds (probe-irrelevant) and accuracy in percent correct (irrelevant-probe) for the unfamiliar-probe test

Condition	Delay		
	10 min	24 h	1 week
RT			
Deep	11.17 (12.91)	−13.08 (5.23)	8.43 (7.15)
Shallow	−10.59 (8.10)	1.96 (9.37)	0.94 (12.13)
Accuracy			
Deep	0.86% (1.0%)	−0.19% (1.0%)	0.34% (0.50%)
Shallow	1.76% (1.5%)	−0.60% (1.0%)	−0.65% (1.0%)

Positive values indicate that probe responses were slower or less accurate than irrelevant responses. Data are shown as a function of study-test delay and probe study condition

Values in parentheses are 1 ± SEM

Typically in this paradigm, familiar-probe responses are both slower and less accurate than on irrelevant trials (Seymour and Kerlin 2008; Seymour et al. 2000). However, in the present study, probe and irrelevant accuracies were near ceiling for both tests leading to an attenuated accuracy effect. Based on debriefing of participants, the $10 reward for "beating the test" was highly motivating and led to cautious responding on probe and target trials. Although participants were able equate probe and irrelevant accuracy rates, they were unable to speed up probe responses sufficiently to attenuate the RT effect. Thus we were able to detect participants' knowledge of the probes phrases, despite their attempts to conceal it. The strong test efficiency in light of the attenuated accuracy effect is promising because the $10 reward for "beating the test" in the current procedure represents a stronger incentive than in previous RT-based CKT studies (Rosenfeld et al. 2004; Seymour and Kerlin 2008; Seymour et al. 2000).

The classification results were striking. For immediate tests, probe elaboration did not influence the test and an overall classification accuracy (familiar-probes and unfamiliar-probes) of 93% was observed. However, after 1 week, overall classification for well-elaborated probes remained high at 90%, whereas classification using poorly elaborated probes approached chance.

General Discussion

The present study is the first to explicitly examine or demonstrate an interaction of time delay and encoding depth on the efficiency of the RT-based CKT. Previous studies have only examined time delay *or* encoding depth. The results show that for well-learned information, detection accuracy was high and relatively stable over the examined time period. Less elaborated items were accurately detected after 10 min, but by 1 week test accuracy was significantly compromised. In previous CKT studies, test accuracy was affected by test delay in some cases (Rosenfeld et al. 1991; Waid et al. 1978, 1981a), but not in others (Carmel et al. 2003). The present data suggest that this discrepancy may have been driven by differences in probe elaboration.

Differentiating Centrality and Encoding Level

The present results clearly suggest that for delayed tests, elaborated items make better probes than unelaborated ones. However, examiners creating CKT test-lists can only reasonably determine which items were central to the crime. Unfortunately, centrality does not guarantee rich encoding. For example, while the most highly emotional and salient aspects of a crime scene (e.g., a weapon) seem

to be encoded more readily and more elaborately than less salient items, this heightened attentional focus to some items can decrease the likelihood that other items are well encoded (including other central items). This *weapon-focus* effect has been demonstrated in the literature (e.g., Loftus et al. 1987) and may ensure that some central crime-scene items can serve as effective probes while other may not. Similarly, there may also be peripheral details that are sufficiently encoded to later serve as effective probes. Baddeley (1978) pointed out that there are many conditions in which initially shallow encoding can lead to a durable memory trace. For example, peripheral information may resist forgetting when it is unusual or particularly vivid, contain associated olfactory information, or has associated physical action (Baddeley 1999). Numerous studies have also found that emotionally valenced information is remembered better after a delay than neutral items. This is true for both recall (e.g., Christianson and Loftus 1991), and recognition (e.g., Comblain et al. 2004; Johanssonet al. 2004).

Furthermore, these studies typically show that negatively valenced stimuli lead to more durable memory traces than positive or neutral items. However, in some cases, positively charged items are remembered better (Matt et al. 1992), and can interact with age (Mather and Carstensen 2003). There is also evidence that arousal, and not emotional valence itself, leads to increased retention of emotional items (Bradleyet al. 1992; Kensinger and Corkin 2003; Kleinsmith and Kaplan 1963). This link with general arousal may offer a mechanism whereby peripheral items are encoded well enough to support later recognition.

Future Directions

One question unanswered by these data is the potential effect of explicit countermeasure instructions on both the immediate and delayed RT-based CKT. Successful countermeasures have been reported that significantly reduce the efficiency of most notable CKT variations (e.g., Ben-Shakhar and Dolev 1996; Elaad 1999; Hontse et al. 1996; Rosenfeld et al. 2004). Although we did not offer participants an explicit countermeasure strategy, the present study does offer some insight. Participants were informed about how the concealed knowledge effect was scored (comparison of the speed and accuracy differences between probe and irrelevant responses) and offered a 10$ financial incentive to equate these responses. The resulting attenuation of the accuracy effect compared to previous RT-based CKT reports suggests that participants were indeed trying to beat the test. However, the classification accuracy found for elaborated probes was similar to Seymour et al. (2000) who gave participants a specific countermeasure strategy, but offered a weaker incentive.

Some researchers has also suggested that due to factors such as level of motivation, laboratory tests may overestimate the size of the concealed knowledge effect examiners can expect in the field (Ben-Shakhar and Elaad 2003; Ben-Shakhar and Furedy 1990; Carmel et al. 2003; Elaad 1990; Gronauet al. 2005). Typically these studies report larger effect sizes in laboratory settings than in the field, although examining whether this difference affects detection efficiency is not as common. A study by Pollina et al. (2004) showed that despite differences in effect sizes, classification accuracy was the same in the laboratory and in the field. Similarly, a large meta-analysis of CKT studies revealed a significant difference in test effect-size when "highly motivated" participants ($d = 1.76$) were compared to those with "low motivation" ($d = 1.34$), but not on their respective test efficiencies ($a = 0.82$ and 0.80, respectively; Ben-Shakhar and Elaad 2003).

Concern for the applied forensic use of the CKT has also been expressed by researchers who examine real crime casefiles (Elaad 1990; Elaad et al. 1992; Podlesny 2003). For example, Podlesny examined 758 FBI case files and found only 15% with critical details sufficiently specific to the case to serve as "good" probes, and only 11% with details known solely to the FBI. The implication of this finding depends on how one intends to use the CKT. If the goal is to identify guilty suspects, then specific probes shielded from leakage are needed. Podlesny's report shows that in many FBI investigations this requirement is not met. However, if the goal is to identify suspects or witnesses, or to eliminate suspects from further consideration (i.e., focus on finding those who have relevant knowledge), then the CKT is an accurate and reliable tool. We note that even one or two "good" probes may be sufficient. Although six or more items are often considered ideal, number of probes does not appear to be highly correlated with CKT effect size or test efficiency (Ben-Shakhar and Elaad 2003).

Finally, although the present interaction between delay and encoding can be logically extrapolated to delays longer than the 1-week period we examined, more research is needed to test longer delays and also to compare multiple CKT variations under such conditions. We also expect the relationship between time and encoding shown here to apply to other CKTs measures.

Conclusion

We've provided the first clear demonstration of how probe elaboration and test delay interact in their influence on CKT accuracy. The present results support test administrators' typical bias towards using more central crime details in applied detection tasks, but they also suggest that for immediate tests, peripheral probes may be just as effective. When these data are considered in light of memory research

that suggests central crime-details can remain unelaborated and peripheral details can be richly encoded, the interpretation of negative test results is no longer straightforward. More research is needed on the interaction between delay and encoding that includes explicit countermeasures and longer delays.

Acknowledgments The authors thank Jess Kerlin, Christina Bolanos, and those who participated in this study. We also thank Colleen Seifert, Eric Schumacher, Chris Baker, and anonymous reviewers for their comments and suggestions.

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*, 504–522.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review, 85*, 139–152.

Baddeley, A. D. (1999). *Essentials of human memory*. Hove, England: Psychology Press.

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

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology, 88*, 131–151.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective* (p. 169). New York, NY, US: Springer.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 379–390.

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: The external validity of mock crime studies. *Journal of Experimental Psychology: Applied, 9*, 261–269.

Christianson, S.-Ã. k., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion, 5*, 81–108.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671–684.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology, 75*, 521–529.

Elaad, E. (1997). Polygraph examiner awareness of crime-relevant information and the guilty knowledge test. *Law and Human Behavior, 21*, 107–120.

Elaad, E. (1999). The challenge of the concealed knowledge polygraph test. *Expert Evidence, 6*, 161–187.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology, 77*, 757–767.

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

Furr, M. R., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2*, 45–67.

Gronau, N., Ben-Shakhar, G., & Cohen, A. (2005). Behavioral and physiological measures in the detection of concealed information. *Journal of Applied Psychology, 90*, 147–158.

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.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (2005). Scientific status: The case for polygraph tests. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (Vol. 4). Minnesota: Thompson West.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*, 1169–1180.

Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology, 65*, 190–193.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, Massachusetts: Harvard University Press.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and Human Behavior, 11*(5), 5–62.

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied, 2*, 164–177.

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

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

Mather, M., & Carstensen, L. L. (2003). Aging and attentional biases for emotional faces. *Psychological Science, 14*, 409–415.

Matt, G. E., Vazquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12*, 227–255.

Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 49–86). New York: Academic Press.

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

Patalano, A. L., & Seifert, C. M. (1994). Memory for impasses during problem solving. *Memory & Cognition, 22*, 234–242.

Podlesny, J. A. (2003). A paucity of operable case facts restricts applicability of the guilty knowledge technique in FBI criminal polygraph examinations. *Forensic Science Communications, 5*, Retrieved March 6, 2006, from http://www.fbi.gov/hq/lab/fsc/backissu/july2003/index.htm.

Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal of Applied Psychology, 89*, 1099–1105.

Raskin, D. C. (Ed.). (1989). *Psychological methods in criminal investigation and evidence*. New York, NY: Springer.

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*, 319–335.

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*, 251–259.

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, 24*, 157–161.

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.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide (Version 1.1)*. Pittsburgh, PA: Psychology Software Tools.

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

Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge". *Journal of Applied Psychology, 85*, 30–37.

Waid, W. M., Orne, E. C., Cook, M. R., & Orne, M. T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of information. *Journal of Applied Psychology, 63*, 728–733.

Waid, W. M., Orne, E. C., & Orne, M. T. (1981a). Selective memory for social information, alertness, and physiological arousal in the detection of deception. *Journal of Applied Psychology, 66*, 224–232.

Waid, W. M., Wilson, S. K., & Orne, M. T. (1981b). Cross-modal physiological effects of electrodermal lability in the detection of deception. *Journal of Personality and Social Psychology, 40*, 1118–1125.