Different Target Modalities Improve the Single Probe Protocol of the Response Time-Based Concealed Information Test

Koller, D.; Hofer, F.; Verschuere, B.

DOI
10.1016/j.jarmac.2021.08.003
Publication date
2022
Document Version
Final published version
Published in
Journal of Applied Research in Memory and Cognition

AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

2022, Vol. 11, No. 1, 135–141
https://doi.org/10.1016/j.jarmac.2021.08.003

EMPIRICAL ARTICLE

Different Target Modalities Improve the Single Probe Protocol of the Response Time-Based Concealed Information Test

Dave Koller[1, 2], Franziska Hofer[3, 4], and Bruno Verschuere[2]
[1] Department of Psychology, University of Zurich
[2] Department of Clinical Psychology, University of Amsterdam
[3] Zurich State Police, Airport Division, Research and Development, Switzerland
[4] Brainability, Developing Human & Organizational Potentials, Zurich, Switzerland

To detect if someone hides specific knowledge (called "probes"), the response time-based Concealed Information Test (RT-CIT) asks the examinee to classify items into two categories (targets/non-targets). Within the non-targets, slower RTs to the probes reveal recognition of concealed information. The preferred protocol examines one piece of information per test block (single probe protocol), but its validity is suboptimal. The aim of this study was to improve the validity of the single probe protocol by presenting the information in multiple modalities. In a preregistered study (N = 388) participants were instructed to try to hide their nationality. The items referring to the nationality were presented as words, flags, and maps. Increasing the number of modalities of the targets (BF_{10} = 37), but not of the probes and irrelevants (BF_{01} = 6), increased the CIT-effect. This broadens the range of the RT-CIT's applicability, which is an important step towards application in practice.

General Audience Summary

Law enforcement agencies, intelligence services, or even companies can encounter situations in which they would like to find out if a person knows something about an incident even though he or she claims not to. In situations like these, the response time-based Concealed Information Test (RT-CIT) could be applied. In this automated test, examinees first learn a few so-called target items (words and/or pictures). Then they will see different items one by one. These items are either the learned targets, items related to the incident (so-called probes), or unrelated items (so-called irrelevant items). The examinees' task is to indicate with two keys on the keyboard if he or she recognizes this item or not (i.e., YES-button for the targets and NO-button for the other items). Slower responding to items related to the incident than to irrelevant items is an indication of concealed recognition. The RT-CIT works best if multiple pieces of information about the incident (e.g., location, stolen jewelry, tool used for the break-in) are tested. In reality, however, an examiner might have only one information that can be tested. We explored two possibilities to improve the performance of the RT-CIT in such a scenario. In an online study with 388 participants, they were asked to hide their true nationality and claim to be from another country which was used as the target information. We investigated if the performance of the RT-CIT could be increased if we present the items not only as words (e.g., "United Kingdom") but also as flags and maps of countries. Presenting the target information in different ways increased the difference between probes and irrelevant items which implies more correct detections. This makes the RT-CIT applicable in a wider range of situations.

Keywords: memory detection, Concealed Information Test, CIT, deception, single probe protocol, lie detection

This article was published Online First October 7, 2021.

We thank the Swiss Federal Office of Civil Aviation (project number: 2016-106) and the Zurich State Police, Airport Division for their financial support.

The authors declare that they have no conflict of interest.

Dave Koller proposed the initial study design which was refined in collaboration with Bruno Verschuere Programming, data collection, and analysis was done by Dave Koller. The manuscript was mainly written by Dave Koller but in close collaboration with Bruno Verschuere and in consultation of Franziska Hofer. Franziska Hofer and Bruno Verschuere supervised the project.

The preregistration, material, data, and scripts that support the findings of this study are openly available on OSF (https://osf.io/d536j/).

Correspondence concerning this article should be addressed to Dave Koller, Department of Psychology, University of Zurich, Zurich, Switzerland. Email: d.koller@psychologie.uzh.ch

The Concealed Information Test (CIT; Lykken, 1959) can be used to reveal if a person has specific knowledge he/she claims not to possess and is frequently used by Japan's police (Osugi, 2018). The rationale of the CIT is that a person shows a different reaction to an item whose recognition he/she tries to conceal compared to similar yet irrelevant items (CIT-effect). When for example crime related items (e.g., the location where the victim was found: the woods) are presented among other plausible items (e.g., the river, the sewer, the shed), only a person with crime knowledge is expected to show a distinct reaction to the crime related items (so-called *probes*) compared to the other items (so-called *irrelevants*). The typically observed reactions for concealed recognition in the CIT include increased response times, increased skin conductance response, and increased P300 amplitude (see Verschuere & Meijer, 2014 for a review). With its potential to easily test large groups of people remotely (Verschuere & Kleinberg, 2016), there has been renewed interest in response times as a CIT index.

The response time-based CIT (RT-CIT) includes a third item type, the so-called targets (Farwell & Donchin, 1991; Rosenfeld et al., 1988; Seymour et al., 2000). Examinees are instructed to press the "YES" button only for targets and the "NO" button for all other items (including the probes). Meijer et al. (2016) found a weighted average of the area under the receiver operating characteristic curve (AUC) of 0.82 based on 981 participants across the nine analysed experiments, which is well above chance. The classification performance is known to vary with several factors, one of which is the testing protocol (Lukács et al., 2017; Verschuere et al., 2015).

There are two main RT-CIT testing protocols: the *single probe protocol* and the *multiple probe protocol*. In the single probe protocol, each block contains items of one item category (Lykken, 1959). For instance, a first block could test the examinee for stolen goods, the next block for crime locations, etc. In the multiple probe protocol, the items of the different item categories are all presented intermixed in each block (e.g., Seymour et al., 2000). Research showed superior classification performance for the multiple probe protocol compared to the single probe protocol in the RT-CIT (Lukács et al., 2017; Verschuere et al., 2015). Experiment 2 of Verschuere et al. (2015), for example, showed larger effect sizes for the multiple probe protocol compared to the single probe protocol ($d_{\text{within, MP}}$ = 1.52; $d_{\text{within, SP}}$ = 0.59) as well better classification (AUC_{MP} = .86; AUC_{SP} = .69). The application of the multiple probe protocol is limited to situations in which more than one critical piece of information is known. Furthermore, when the RT-CIT is used as an investigative tool in the form of a searching RT-CIT (Koller et al., 2020), the single probe protocol may be the only option. For example, to reveal the exact location of a planned terror attack, the examiner would first need to test for the city, then for a specific location within that city. Without this serial testing, the number of items that would need to be included in the CIT would increase rapidly.

Given the need for a more accurate single probe protocol, Lukács et al. (2017) introduced the addition of familiarity related filler items that needed to be classified as familiar (e.g., the filler word "RECOGNIZED") or unfamiliar (e.g., the filler word "UNKNOWN"). This modification led to larger CIT-effects as it assures semantic processing of the stimuli and/or may enhance response conflict for the probes. While familiarity related fillers seem to be a good way to improve the single probe protocol, exploring alternative solutions still has its merits (e.g., to find combinations of effective techniques or to overcome potential shortcomings of one solution).

In the present study, we examined whether presenting items in different modalities (e.g., a country as name, flag, or map) is sufficient to increase the validity of the single probe protocol.[1] Further, we explored whether it is the probe or the target modalities that contribute to the effect. The ultimate goal of this study was to make the RT-CIT applicable to a wider range of scenarios.

Method

The experiment was approved by the ethical committee of the Faculty of Social and Behavioural Sciences of the University of Amsterdam (approval number: 2014-CP-3389). Preregistration, material, data, and scripts can be found on https://osf.io/d536j/.

Deviations From Preregistration

We had three exclusion criteria, one stated: "Participants with mean RT of irrelevant items deviating more than ± 3 *SE* of their respective group means of irrelevant items (only correctly answered trials are considered in this analysis) will be excluded." However, this is an unfortunate typing error and was meant to state "± 3 *SD*."

Participants

Participants were eligible to enrol if they were between 18 and 45 years old and of one of the following nationalities: British, Portuguese, Spanish, German, Italian, Austrian, or Swedish (see Procedure). Completion of this study took participants about 14 min and was reimbursed with 1.4 GBP (≈ 1.8 USD).

Following the preregistered recruitment procedure, 400 participants were recruited using the online platform Prolific (M_{age} = 27.88, *SD* = 7.36, 51% female). Twelve participants (3%) were excluded based on the three preregistered criteria. Eight indicated that they provided wrong information about their nationality in the pre-CIT questionnaire, three due to poor performance in the task (more than 50% errors in at least one item type), and one due to slow RTs to irrelevant items ($M\ \text{RT}_{\text{irrel}}$ > group mean + 3**SD*), resulting in a final sample of $N = 388$ (M_{age} = 27.74, *SD* = 7.35, 51% female). Per inclusion criteria, participants were British (42%), Portuguese (33%), Spanish (11%), German (2%), Italian (10%), Austrian (0%), and Swedish (2%).

Procedure

At the beginning of the experiment, participants were asked to indicate their nationality from a list of the seven nationalities that

[1] We infer increased validity from larger within-participant CIT-effects. Lukacs and Specker (2020) showed that this inference might not be valid if the standard deviations of the within-participant CIT-effect increases substantially. This was not the case in our study (see Table 3 and supplementary materials on OSF; https://osf.io/d536j/).

were eligible for the study and "another country."[2] Participants that chose "another country" were directed to the end of the study, explaining that they are not eligible to participate in this study. The chosen country was the critical probe item in the CIT. We used Prolific's built-in nationality filter to invite eligible participants only. To further improve data quality, we asked participants at the end of the study whether they indicated their true nationality in the beginning. It was made clear that this does not have any influence on their payment or on in-/exclusions for future studies. These measures give us confidence in the truthfulness of the reported nationalities.

If participants chose one of the seven countries as their nationality, they were subsequently asked to indicate up to one other country from the remaining six that is of significance to them. This country was removed from the item pool of irrelevant items in the CIT to assure that only the country of origin stands out amongst the other countries in the CIT. Not removing other significant countries would lead to a lower sensitivity. If no other country was indicated as significant, one was removed at random.

Next, participants were asked to imagine a scenario in which an online service they need to use is not available for people of certain nationalities, and theirs is one of them (e.g., Sweden). Therefore, they had to pretend to be from another country (randomly assigned from the remaining five countries, e.g., Spain). As they were suspected of cheating (the online service provider was said to have detected a mismatch between the location of the computer's IP address and the claimed nationality), they were asked for additional verification: the RT-CIT. Participants were told that this verification tries to detect their true nationality and that their goal is to hide that information and to convince the service provider to be from their indicated country (Spain). Then, the RT-CIT started. After the RT-CIT, participants were asked if they indicated their true nationality at the beginning of the study, thanked, and redirected to the Prolific website. The payment was processed through Prolific within a few days.

RT-CIT

The RT-CIT was programmed with Inquisit 5 (2016) and ran on the participants' computer using the Inquisit Web plugin, which they downloaded just before starting the test.

Participants were instructed to answer the question "Is this your nationality?" as fast and accurately as possible by responding YES (A-key) only for items that belong to their fake nationality (*targets*) and NO (L-key) to all other items (*probes* and *irrelevants*) and therefore denying their true nationality. So, a person from Sweden pretending to be from Spain would answer YES only to Spain, and NO to Austria, *Sweden*, Portugal, and Germany. Three modalities (word, flag, map), with six items (1 probe, 1 target, 4 irrelevant) per modality were used as items in the RT-CIT. The three target modalities (i.e., word, flag, map corresponding to their fake identity) were presented on the screen for 10 s together with a reminder that participants should respond to these items with YES, followed by a repeatable practice block in which each of the 18 items was presented once. Then followed three test blocks, each consisting of six burn-in trials at the beginning of the block that were discarded to avoid possible artefacts (e.g., due to finger placement, accustoming to the pace) and 126 test trials (totalling 378 test trials).

Every trial started with the stimulus presentation in the centre of the screen which stayed until a response was given or the response deadline of 1.5 s was reached. The question "Is this your nationality?" was displayed on top of the screen as well as the answer labels (YES, NO) and key labels (A-key, L-key) to the left and right of the stimuli, see Figure 1. Feedback (red "WRONG" displayed below the stimulus for 250 ms for incorrect responses; red "TOO SLOW" above the stimulus 800 ms after stimulus onset) was given only in the practice trials. The practice block was repeated up to three times or until the accuracy of each item type (probe, target, irrelevant) was above 50% and the mean RT of irrelevant items was below 800 ms. Reminder instructions were given before each repetition. The response-stimulus-interval varied randomly from 400 to 800 ms. Until the end of the practice phase, the task was identical for all participants.

For the test phase, we manipulated the number of modalities of targets (one vs. three) and the number of modalities of probes and irrelevant items (one vs. three) that participants saw in any given block. Participants had been randomly divided into four groups (1Probe-1Target with $n = 89$, 3Probe-1Target with $n = 103$, 1Probe-3Target with $n = 96$, 3Probe-3Target with $n = 100$) based on the automatically generated participant number. The 1Probe-1Target modality condition (see Figure 1a) used one probe and one target of the same modality in each block (e.g., flags in the first, words in the second, and maps in the third block). This condition mimics the traditional single probe protocol. The 3Probe-1Target modality condition (see Figure 1b) used all three modalities of the probe but only one modality of the target per block. This means that in one block, the target information was only presented as a flag, for example, while the probes and irrelevants were presented in all three modalities. Similarly, the 1Probe-3Target modality condition (see Figure 1c) used one modality for the probes and irrelevants but three modalities for targets per block. The 3Probe-3Target modality condition (see Figure 1d) used all modalities for probes, irrelevants, and targets which results in three identical blocks. Table 1 shows the word-block for all conditions as an example. Each participant saw every item 21 times over the course of the three blocks. The blocks were presented in random order.

Results

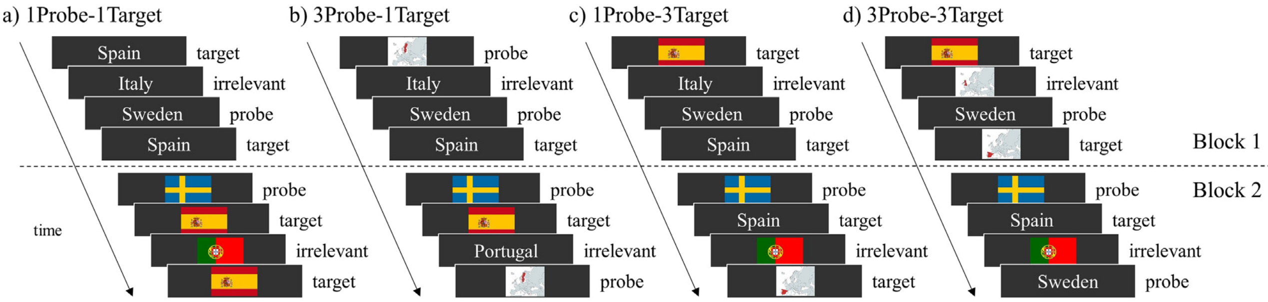
We only included correctly answered non-target test trials with RTs between 200 ms and 1500 ms in the analysis. A total of 1209 non-target trials (0.99%) were excluded.[3] We obtained the typical

[2] Due to a programming mistake, participants did not see the informed consent form and therefore did not give us explicit consent before the study. However, in agreement with the ethics review board of the University of Amsterdam, we are convinced that it is ethical to use the collected data for the following reasons. 1) Data collection was done on Prolific, an online platform for running scientific studies to which participants signed up. 2) Participants were informed about the nature of the study before they chose to participate. More specifically they were informed that they need to overcome an information verification test, that the test is based on reaction times, that the test cannot be paused, that they need to install a plugin but instructions to uninstall the plugin will be provided, that only participants of certain nationalities are eligible to participate (UK, Germany, Spain, Portugal, Italy, Sweden, Austria), and they were informed about the time the study will take and reimbursement they will receive. 3) Participants had the possibility to revoke their consent at any point by "returning" their submission, as Prolific calls it. 4) Participants could revoke their consent by contacting the first author using Prolific's built-in messaging system.

[3] Due to the low error rate and following the preregistration, errors rates were not analysed. Error rates per condition are presented in the supplementary materials on OSF (https://osf.io/d536j/).

Figure 1

Exemplary Segments of the Four Conditions



Note. This illustration shows exemplary segments from two of the three blocks for each condition. See the online article for the color version of this figure.

CIT-effect in response times (i.e., larger RTs for probes than for irrelevant items) in each of the four conditions, see Table 2. Consequently, the main dependent variable, the within-participant CIT-effect ($dCIT = (M\ RT_{Probe} - M_{Irrel})/SD\ RT_{Irrel}$; Kleinberg & Verschuere, 2015), was credibly greater than zero for all groups (i.e., the lower bound of the 95% CI greater than zero; see Table 2).

The *dCIT* reported is the standardized within-participant probe-irrelevant difference and is not to be confused with Cohen's *d*. The effect size δ and 95% credible intervals were estimated with the Bayesian Wilcoxon signed-rank test of JASP (JASP Team, 2020) using a Cauchy prior (scale = 0.707) because the normality assumption needed to calculate Cohen's *d* was violated.

We used the BayesFactor package (Morey & Rouder, 2018) extension for R (R Core Team, 2020) to conduct a Bayesian fixed effects ANOVA with a JZS prior (Cauchy prior with scale = 0.5) to test for effects of number of probe modalities and number of target modalities on the within-participant CIT-effect (*dCIT*). The data was most likely under M_{Tar}, the model containing only the main effect of number of target modalities (Table 3). We found very strong evidence for this model compared to the null-model M_0 ($BF_{Tar,\,0}$ = 37.1), supporting the hypothesis that there is an effect of number of target modalities on the within-participant CIT-effect (*dCIT*). Both model comparisons that can be used to assess the effect of number of probe modalities showed moderate evidence against such an effect ($BF_{Pro,\,0} = 0.17$; $BF_{Tar,\,Main} = 5.3$). Those comparisons show that the data is about 5 times more likely under the model without an effect of number of probe modalities. We also found evidence against an interaction effect ($BF_{Main,\,Full} = 6.3$; $BF_{Tar,\,Full} = 33.5$).[4] Therefore, our results showed a benefit of using multiple target modalities but no effect of multiple probe modalities.

The magnitude of the effect of target modality was assessed by the parameter's (β_{Tar}) posterior distribution of M_{Tar} ($M\ \beta_{Tar} = 0.10$; 95% HDI = [0.04, 0.15]). Because the 95% HDI does not include 0, it can be concluded that presenting the targets in three different modalities instead of one lead to a credible increase of the within-participant CIT-effect (*dCIT*) of 0.1, on average. This effect is independent of the number of probe modalities, since the model comparison showed evidence against an interaction effect.

Discussion

Some situations in practice require the single probe protocol of the RT-CIT that tests for one item per block, but it has lower validity than the multiple probe protocol that tests several items per block. The present study aimed at investigating how the validity of the single probe protocol could be improved to make it applicable in a wider range of situations. We tested for one piece of information (nationality) and presented it either in one or in three different modalities (word, flag, map). We independently manipulated the number of target modalities and probe modalities. We found moderate evidence against an effect of the number of probe modalities on the within-participant CIT-effect (*dCIT*) but strong evidence for an effect of the number of target modalities. This study suggests that the validity of the single probe protocol can be increased not by presenting the probes in different modalities but by presenting the targets in different modalities.

The impact of the number of target modalities is especially interesting, as researchers initially thought targets would not matter at all and could even be discarded (e.g., Matsuda et al., 2009). Nevertheless, it has become clear that the target items influence the CIT-effect. Gamer et al. (2017) argued that perceptual similarity between target items and test items (probes and irrelevants) influences the encoding of test items. Dissimilar target items can be identified easily without deep encoding, leading to smaller probe-irrelevant differences. However, not only perceptual similarity seems to impact the CIT-effect. Suchotzki et al. (2018) manipulated the familiarity of the target items. They argued that this increased the feature overlap (in the familiarity feature) between targets and probes, and therefore the response conflict for probes, which lead to larger probe-irrelevant differences. Suchotzki et al. (2018) also observed a small increase in the probe-irrelevant difference when four targets were used as compared to two targets. A crucial difference to our study is that they added targets (e.g., "Spain" and "Greece") whereas we presented the same target in different modalities (e.g., flag and name of Spain). This increased the number of semantic objects participants had to keep in mind.

The reason why we found larger CIT-effects in the three target modality conditions could have been because it might have altered the way examinees approached the task. With a single target modality, examinees can perform the task by focussing on a unique perceptual feature of the target (while attempting to ignore other features). Such perceptual processing reduces the influence of other

[4] A reviewer proposed an alternative model comparison (using Baws factors), which essentially lead to the same results. This analysis can be found in the supplementary materials on OSF (https://osf.io/d536j/).

Table 1
Experimental Conditions Illustrated for the Word-Block

	Item type		
	Probe	Target	Irrelevant
1P1T	Sweden	Spain	Germany, Portugal, Austria, Italy
3P1T	Sweden,  	Spain	Germany,  
1P3T	Sweden	Spain,  	Germany, Portugal, Austria, Italy
3P3T	Sweden,  	Spain,  	Germany,  

Note. Cells that only contain words would be changed to flags or maps for the flag-block and map-block, respectively. The mixed cells remain unchanged. See the online article for the color version of this table.

features needed for the CIT-effect (i.e., familiarity, saliency). Targets share those features with the probes but not the irrelevants, which leads to slower RTs for probes than for irrelevants (Gamer et al., 2007; Suchotzki, 2018). With multiple target modalities, there is not a single perceptual feature that allows doing the task and therefore semantic encoding is required. This leads to the incorporation of a wider array of features in the decision-making process, including features that lead to response conflict for probes. We call this explanation *target focus hypothesis*. A large body of research on the Stroop effect (Stroop, 1935) and the Garner interference (Garner, 1974) showed that it is possible to primarily focus the attention on the feature of interest but other, irrelevant, features are often not ignored completely and thus influence the decision also (for reviews, see Algom & Fitousi, 2016; MacLeod, 1991). Therefore, two crucial processes for the target focus hypothesis, focused attention on specific features and integration of information from multiple feature dimensions into a binary decision, have been shown in other paradigms. Note that a connection between Stroop-like interference and the CIT-effect has been suggested before (e.g., Seymour & Schumacher, 2009). The addition of other ways to increase the reliance on these conflict inducing features, as it was done by using familiar targets (Suchotzki et al., 2018) or familiarity-related filler items (Lukács et al., 2017) could improve our multiple modality single probe protocol even further.

It is fortunate, especially for applied purposes, that it seems to be sufficient to increase the target modalities rather than the probe modalities. It would be an additional restriction if only probes could be used for which different presentations not only exist, but for which the examinee also has a strong internal representation—an important factor to obtain strong CIT-effects (Geven et al., 2019). For example, if the police want to test someone only on the pseudonym of a cybercriminal, it would be very challenging to find different visual modalities for that pseudonym.

Limitations and Future Studies

This study exclusively looked at nationality information and limited the number of countries that were included in the RT-CIT. Of those seven countries, 75% of participants were British or Portuguese. While we expect that our results generalize to other countries, a replication study with a more diverse population and a balanced design should be conducted.

Due to the randomization and the nature of an online study, we cannot rule out some degree of selective attrition. However, it seems unlikely that it is a major issue since the group sizes were very similar.

Furthermore, we cannot exclude that our results are unique to the modalities (word, flag, map) used in this study. While words are

Table 2
Response Times, Within-Participant CIT-Effect, and Probe-Irrelevant Effect Size

	RT				
Condition	Probe	Irrelevant	Target	*dCIT*	δ
1P1T	418 (64)	405 (52)	491 (51)	0.15 (0.30)	0.61 [0.38, 0.85]
(*n* = 89)				[0.10, 0.23]	
3P1T	458 (58)	441 (51)	520 (51)	0.17 (0.24)	0.82 [0.59, 1.04]
(*n* = 103)				[0.13, 0.22]	
1P3T	458 (57)	436 (51)	545 (54)	0.24 (0.28)	1.08 [0.75, 1.39]
(*n* = 96)				[0.19,0.30]	
3P3T	509 (59)	482 (55)	575 (53)	0.28 (0.32)	1.03 [0.75, 1.27]
(*n* = 100)				[0.23,0.35]	

Note. CIT = Concealed Information Test; RT = response time. Mean response times (in ms; *SD* in parentheses), mean within-participant CIT-effect (*SD* in parentheses; 95% credible interval in brackets), and probe-irrelevant effect size δ (95% credible interval in brackets) estimated with a Bayesian Wilcoxon signed-rank test.

Table 3
Bayesian Model Comparison

Models		Compared to the intercept only model (i.e., $BF_{Model,\ 0}$)	Compared to the best model (i.e., $BF_{Tar,\ Model}$)
M_{Pro}	*dCIT* ~ nProbes	0.17 (±0%)	213.38 (±0%)
M_{Tar}	*dCIT* ~ nTargets	37.12 (±0%)	1.0 (±0%)
M_{Main}	*dCIT* ~ nProbes + nTargets	7.06 (±0.97%)	5.26 (±0.97%)
M_{Full}	*dCIT* ~ nProbes + nTargets + nProbes × nTargets	1.11 (±1.06%)	33.46 (±1.06%)

Note. CIT = Concealed Information Test. In the first column, the models are compared to the null-model. Bayes factors >1 indicate that the data is more likely under this model than the null-model. In the second column, the best model (M_{Tar}) is compared to the other models. Bayes factors indicate how much more likely the data is under M_{Tar} compared to the other models.

commonly used when testing for participants' nationality, flags and maps have not been studied often with the RT-CIT. Exploring the possibilities of switching between visual and auditory presentation within a block or other means that prevent participants to focus on a perceptual feature (e.g., using synonyms or very closely related stimuli as targets [e.g., gun, pistol, firearm]) seems highly relevant for applied purposes.

An obvious but important limitation to consider from an applied perspective is that the increased validity for multiple target modalities might not generalize from the RT-CIT to the physiological CIT, the only version currently applied in the field (Osugi, 2018). It is likely that different mechanisms are involved in the RT-CIT and the physiological CIT (see e.g., Klein Selle et al., 2018; Seymour & Schumacher, 2009).

With the introduction of familiarity related fillers (Lukács et al., 2017) and our multiple modality approach, we now know of two ways to increase the validity of the single probe protocol. We want to encourage future research to explore the combination of both approaches.

Conclusion

Our findings show that in situations in which only one piece of information is available for testing, the validity of the RT-CIT can be increased by using multiple modalities for the target item. Presenting the target in several modalities may prevent a purely perceptual way of processing and assure semantic processing of the stimuli in the RT-CIT. This brings about the processing of feature dimensions that induce response conflict and therefore the CIT-effect.

References

Algom, D., & Fitousi, D. (2016). Half a century of research on garner interference and the separability-integrality distinction. *Psychological Bulletin*, *142*(12), 1352–1383. https://doi.org/10.1037/bul0000072.

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

Gamer, M., Bauermann, T., Stoeter, P., & Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Human Brain Mapping*, *28*(12), 1287–1301. https://doi.org/10.1002/hbm.20343.

Garner, W. R. (1974). *The processing of information and structure*. Erlbaum.

Geven, L. M., Ben-Shakhar, G., Kindt, M., & Verschuere, B. (2019). It's a match!? Appropriate item selection in the Concealed Information Test. *Cognitive Research: Principles and Implications*, *4*, 11. https://doi.org/10.1186/s41235-019-0161-8.

Inquisit. (2016). *Inquisit* (Version 5) [Computer software]. https://www.millisecond.com/.

JASP Team. (2020). *JASP (Version 0.14.0.0)* [Computer software]. https://jasp-stats.org/.

Klein Selle, N., Verschuere, B., & Ben-Shakhar, B. (2018). Concealed information test: Theoretical background. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception* (pp. 35–57). Academic Press. https://doi.org/10.1016/B978-0-12-812729-2.00002-1.

Kleinberg, B., & Verschuere, B. (2015). Memory detection 2.0: The first web-based memory detection test memory. *PLoS ONE*, *10*(4), e0118715. https://doi.org/10.1371/journal.pone.0118715.

Koller, D., Hofer, F., Grolig, T., Ghelfi, S., & Verschuere, B. (2020). What are you hiding? Initial validation of the reaction time-based searching concealed information test. *Applied Cognitive Psychology*, *34*(6), 1406–1418. https://doi.org/10.1002/acp.3717.

Lukács, G., Kleinberg, B., & Verschuere, B. (2017). Familiarity-related fillers improve the validity of reaction time-based memory detection. *Journal of Applied Research in Memory and Cognition*, *6*(3), 295–305. https://doi.org/10.1016/j.jarmac.2017.01.013.

Lukács, G., & Specker, E. (2020). Dispersion matters: Diagnostics and control data computer simulation in Concealed Information Test studies. *PLoS ONE*, *15*(10), e0240259. https://doi.org/10.1371/journal.pone.0240259.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388. https://doi.org/10.1037/h0046060.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Review*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163.

Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., & Takasawa, N. (2009). Event-related brain potentials during the standard autonomic-based concealed information test. *International Journal of Psychophysiology*, *74*(1), 58–68. https://doi.org/10.1016/j.ijpsycho.2009.07.004.

Morey, R. D., & Rouder, J. N. (2018). *Computation of Bayes Factors for common designs (Version 4.2)* [Computer software]. https://richarddmorey.github.io/BayesFactor/.

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology*, *53*(5), 593–604. https://doi.org/10.1111/psyp.12609.

Osugi, A. (2018). Field findings from the Concealed Information Test in Japan. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception* (pp. 97–121). Academic Press. https://doi.org/10.1016/B978-0-12-812729-2.00005-7.

R Core Team. (2020). *R: A language and environment for statistical computing*. [Computer software]. https://www.R-project.org/.

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. https://doi.org/10.3109/00207458808985770.

Seymour, T. L., & Schumacher, E. H. (2009). Electromyographic evidence for response conflict in the exclude recognition task. *Cognitive, Affective, and Behavioral Neuroscience*, *9*(1), 71–82. https://doi.org/10.3758/CABN.9.1.7.

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*(1), 30–37. https://doi.org/10.1037/0021-9010.85.1.30.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651.

Suchotzki, K., De Houwer, J., Kleinberg, B., & Verschuere, B. (2018). Using more different and more familiar targets improves the detection of concealed information. *Acta Psychologica*, *185*, 65–71. https://doi.org/10.1016/j.actpsy.2018.01.010.

Verschuere, B., & Kleinberg, B. (2016). ID-check: Online concealed information test reveals true identity. *Journal of Forensic Sciences*, *61*(S1), 237–240. https://doi.org/10.1111/1556-4029.12960.

Verschuere, B., Kleinberg, B., & Theocharidou, K. (2015). RT-based memory detection: Item saliency effects in the single-probe and the multiple-probe protocol. *Journal of Applied Research in Memory and Cognition*, *4*(1), 59–65. https://doi.org/10.1016/j.jarmac.2015.01.001.

Verschuere, B., & Meijer, E. H. (2014). What's on your mind? Recent advances in memory detection using the concealed information test. *European Psychologist*, *19*(3), 162–171. https://doi.org/10.1027/1016-9040/a000194.

Received December 3, 2020
Revision received August 19, 2021
Accepted August 19, 2021 ■