Received: 13 December 2018 | Revised: 23 April 2019 | Accepted: 25 April 2019

DOI: 10.1002/acp.3565

RESEARCH ARTICLE

WILEY

Information leakage in the Response Time-Based Concealed Information Test

Gáspár Lukács | Ulrich Ansorge

Department of Basic Psychological Research and Research Methods, University of Vienna, Vienna, Austria

Correspondence
Gáspár Lukács, Faculty of Psychology, University of Vienna, Liebiggasse 5, Vienna A-1010, Austria.
Email: gaspar.lukacs@univie.ac.at

Funding information
Austrian Academy of Sciences, Grant/Award Number: 24945

Summary

The Response Time-Based Concealed Information Test (RT-CIT) can reveal when a person recognizes a relevant (*probe*) item among other, irrelevant items, based on comparatively slow responses to the probe item. For example, if a person is concealing his or her true identity, one can use the suspected identity details as probes, and other, random details as irrelevants. However, in our study, we show that even when participants are merely informed about such probes (i.e., the relevant identity details) before performing the RT-CIT, their responses will also be slower to these details. Hence, it is more difficult to distinguish such innocent but pre-informed persons from actually guilty persons. At the same time, we introduce a CIT version with familiarity-related inducer stimuli, but with no targets, that elicits probe-minus-irrelevant RT differences only among guilty participants but not among informed innocent participants. Implications for the theory and the application of CITs are discussed.

KEYWORDS

association, Concealed Information Test, deception, recognition, response time

1 | INTRODUCTION

Undetected deception may have extremely high costs in certain scenarios, such as counterterrorism, pre-employment screening for intelligence agencies, or high-stake criminal proceedings. However, meta-analyses have repeatedly shown that without special aid, based on their own best judgment only, people (including police officers, detectives, and professional judges) distinguish lies from truth on a level hardly better than mere chance (Bond & DePaulo, 2006, 2008; Hartwig & Bond, 2011; Kraut, 1980).

One of the potential technological aids to overcome this problem is the Concealed Information Test (CIT; Lykken, 1959; Meijer, klein Selle, Elber, & Ben-Shakhar, 2014). The CIT aims to disclose whether examinees recognize certain relevant items, such as a weapon used in a recent robbery, among a set of other objects, when the examinees actually try to conceal any knowledge about the criminal case. The recognition of a relevant item can be detected by various means, for instance, by relatively slower responding to relevant items as assessed with a response time-based CIT (RT-CIT). However, the applicability of this test is limited in real-life settings, because it cannot be validly used when an innocent person might also recognize the incriminating item, for example, due to information leakage and the consequential discernible relevance of the critical item (Bradley, Barefoot, & Arsenault, 2011; Podlesny, 2003).

It has been shown in experimental settings that, as expected, innocent participants (i.e., those simulating an innocent criminal suspect) will be more likely to be falsely classified as guilty when informed of the probe, in case of using polygraph or EEG measures (Bradley et al., 2011; Gamer & Berti, 2010; Meijer, Smulders, & Wolf, 2009). However, although theoretically assumed (e.g., Lukács, Gula, Szegedi-Hallgató, &

Csifcsák, 2017), the same effect has yet to be tested using the RT-CIT. Therefore, in the present research, we test this scenario in an information leakage simulation. Furthermore, we introduce a slightly modified RT-CIT method that is resistant to such information leakage. At the same time, these investigations also serve to empirically support an extended theoretical framework of the CIT (Lukács, Grządziel, Kempkes, & Ansorge, 2019; Lukács, Gula, et al., 2017).

1.1 | Three versions of the Response Time-Based Concealed Information Test

The standard RT-CIT consists of a two-alternative forced choice task, where participants classify the presented stimuli as the target or as one of several nontargets by pressing one of two keys (Varga, Visu-Petra, Miclea, & Buş, 2014; Verschuere, Suchotzki, & Debey, 2015). Per each trial, a stimulus is shown. Across trials, typically around five nontargets are presented, among which one is the *probe*, which is an item that only a guilty person would recognize, and the rest are *irrelevants*, which are similar to the probe and thus indistinguishable from the probe for an innocent person. For example, in a murder case where the true murder weapon was a knife, the probe could be the word "knife," whereas irrelevants could be "gun," "rope," and so forth. Assuming that the innocent examinees are not informed about how the murder was committed, they would not know which of the items is the probe. The items are repeatedly shown in a random sequence, and all of them have to be responded to with the same response keys, except one arbitrary *target*—a randomly selected, originally also irrelevant item that has to be responded to with the other response key. Because guilty examinees recognize the probe as the relevant item in respect of the deception detection scenario, it will become unique among the irrelevants and in this respect more similar to the rarely occurring target (Lukács et al., 2016; Lukács, Gula, et al., 2017). Due to this conflict between the instructed response classification of probes as irrelevants on the one hand, and the uniqueness of probes and, thus, greater similarity to the alternative response classification as potential targets on the other hand, the responses to the probes will involve response conflict (Seymour & Schumacher, 2009) and will be generally slower in comparison with the irrelevants. Thus, based on the probe-minus-irrelevant RT differences, guilty examinees can be distinguished from innocent examinees.

A recent study significantly improved this method (i.e., significantly increased the accuracy of distinguishing guilty examinees from innocent ones) by adding inducer items to the task (Lukács, 2019; Lukács, Kleinberg, & Verschuere, 2017), inspired by the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; particularly Bluemke & Friese, 2008; Karpinski & Steinman, 2006; see also Agosta & Sartori, 2013). The IAT measures the strength of associations between certain critical items to be evaluated, such as concepts or entities (e.g., various political parties) and certain attribute items (e.g., good vs. bad). The main idea is that responding is easier (and thus faster) when closely related items share the same response key (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Nosek, Greenwald, & Banaji, 2007). For example (taken from Bluemke & Friese, 2008), a person with an implicit preference for a specific political party responds faster when having to categorize stimuli related to that party (e.g., party emblems or names of well-known party members) together with positive words (e.g., joy and health). Inversely, the categorization of the same stimuli (for the preferred party) will be slower when they share a response key with negative words (e.g., pain and disease).

Note that the general adverse effect of feature overlap (semantic or any other) on categorization is not a novelty of the IAT. In particular, it has long been argued and widely demonstrated that categorization is most efficient in case of "most attributes common to members of the category and the least attributes shared with members of other categories" (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, p. 1435; see also, e.g., Iordan, Greene, Beck, & Fei-Fei, 2015). This, of course, holds for not only concepts but also simple visual stimuli as well (e.g., Azizian, Freitas, Watson, & Squires, 2006; Marchand, Inglis-Assaff, & Lefebvre, 2013).

All in all, we assumed that an analogous mechanism may be introduced in the CIT by adding probe-referring "attributes." (We call these attributes *inducers*, as they serve to induce associations.) In the original study (Lukács, Kleinberg, & Verschuere, 2017), the probes were general autobiographical details (birthday, favorite color, etc.), and correspondingly, the inducers were familiarity- and ownership-related, or, more precisely, self-referring and other-referring. Inducers referring to the participants' own details (e.g., "FAMILIAR" and "MINE") had to be categorized with the same key as the target and, thus, with the key opposite to the response key for the probe (and the irrelevants), whereas inducers referring to other details (e.g., "OTHER" and "THEIRS") had to be categorized with the same key as the probe (and irrelevants). It was assumed that this would have a similar effect as in the IAT: Reponses to the self-related probes (true identity details) would be even slower because they have to be categorized *together* with other-referring expressions (and opposite to self-referring expressions). In contrast, in case of innocents, the probes are not self-related. Hence, the inducers will not slow down the responses to the probe.

Less relevant to the present Introduction, we briefly note that the other additional hypothesized reason for the enhanced effect was that the increased cognitive load (due to the increased complexity) also requires more attention throughout the task, which likely facilitates deeper processing of the stimuli (Lukács, Kleinberg, & Verschuere, 2017, p. 3; see also Visu-Petra, Varga, Miclea, & Visu-Petra, 2013).

In the present study, our first main objective was to test the effect of information leakage on this enhanced version (from here on: *Enhanced-CIT; E-CIT*). For a basis of comparison for the effect of inducers in respect of the information leakage, we also included the original, standard version with no inducers and only a target along with the probe and irrelevants (*Target-CIT*). Although we expected an effect of information leakage (i.e., probe-minus-irrelevant difference for informed innocents) for both versions, the Target-CIT may be less susceptible to this manipulation, simply because it has only a small effect (relatively small probe-minus-irrelevant differences) in case of truly guilty participants in the first place (Lukács, Kleinberg, & Verschuere, 2017; Verschuere, Kleinberg, & Theocharidou, 2015).

Note that here we first used the single-probe protocol CIT, in which only one probe is included within each block of the task (Verschuere & Kleinberg, 2015). We used the multiple-probe protocol CIT (with multiple probes intermixed within each block) in a follow-up experiment, where the relevant differences between the two protocols are also briefly discussed.

Already presupposing that the leakage would indeed render both these versions ineffective, our second main objective was to introduce a leakage-resistant version by a very simple alteration of the E-CIT: removing the target from the task and thereby only leaving inducers along with the probe and irrelevants (*Inducer-CIT*). Our hypothesis here is that response conflict due to the mere recognition of a probe as a relevant item is brought on by the presence of the target.

The target shares the semantic category of the irrelevant and probe items (e.g., dates and in case of looking for a birthday), and its only distinction is that it is the single item that requires a different key response, which makes it unique among the rest of the items, and, consequently, a relevant item in the task. The only other unique and relevant item in the task is the probe. Note that this relevance would be of different origin depending on whether viewed from the perspective of a guilty person or from the perspective of an innocent but informed person. The guilty persons recognize the item as directly related to them—for example, via the committed crime, or because it is their autobiographical detail—whereas the innocent persons recognize the item as one of which they have been informed as relevant to the deception detection test. Nonetheless, in either case, the probe will be recognized as a single relevant item among the irrelevants. Hence, the probe, as opposed to the irrelevants, will share the target's feature of uniqueness and relevance and will therefore be more difficult to categorize together with the irrelevants.

Importantly, if we remove the target from the CIT, there is no such response conflict. Let us consider an example where we try to show whether a person's true country of origin is Germany. A target may be, say, Sweden. In this case, whenever a country appears, the examinee has to consider whether or not it is the target country Sweden, which would require a different key response. However, whenever "Germany" appears, because it is known to the examinee as a relevant country (as suspected country of origin, regardless of whether or not this is true) and therefore unique among the irrelevants, it will take more time to decide that its specific relevance and uniqueness do not invite a different key response. If we do not include the target "Sweden," there is no unique, task-relevant country to be categorized with a different response key. Therefore, relevant or not, all countries, including Germany, will be categorized with the same response keys with equal ease. Neither guilty nor informed innocent participants will have a response conflict due to the relevance of the probe, hence no slower responses to the probe and no probe-minus-irrelevant difference.

However, even without targets, the Inducer-CIT would still be sensitive to self-related information of guilty participants because, for the guilty participants, the probes and the self-referring inducers (that are categorized opposite to the probe) share the feature of self-relatedness. There is, at the same time, no unique, specifically relevant item among the inducers. The inducers are also distinct from the categories of the rest of the items (e.g., countries, to which both probe and irrelevants belong, as well as the target, in the standard CIT). They constitute an additional category of familiarity- and ownership-related words, including two subcategories: one with inducers that refer directly to self-relatedness (and familiarity and ownership) and one with inducers that refer to the opposite (other-relatedness, unfamiliarity, etc.). Probes have to be categorized together with other-related inducers and opposite to self-related inducers. The guilty participant's (but not the informed innocent participant's) relation to the probe is one of self-relatedness, and therefore, in this case, the probe's required response key is in conflict with the key that has to be pressed when self-referring inducers are displayed. Due to this conflict, guilty participants are expected to respond slower to the probe than to the irrelevants (and consequently have larger probe-minus-irrelevant differences). However, in case of innocent participants, being informed of the probe leads only to its recognition as relevant, but not to its association with self-relatedness. Again, note also that there are no unique, relevant items in this task apart from this probe, but merely inducers that create the semantic dimension of self- and other-relatedness and respective associations with response keys. Hence, no conflict, no response slowing to the probes and no probe-minus-irrelevant differences.

For a procedural overview of the three versions, see Table 1.

1.2 | Study overview

In six groups of participants, we tested all three RT-CIT versions (Target-CIT, E-CIT, Inducer-CIT), with two conditions (groups) in each: simple guilty participants (with their own details as probes) and informed innocent participants (with randomly chosen, originally irrelevant details as probes, but thoroughly informed about these details).

2 | EXPERIMENT 1

2.1 | Methods

The experiment was preregistered at https://osf.io/fh6at (Foster & Deardorff, 2017; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). All behavioral data (original and aggregated) are available via https://osf.io/zr39m/, along with the entire original task (written in JavaScript) and its simplified version for demonstration

TABLE 1 Item types in the three task versions

	Target-CIT	E-CIT	Inducer-CIT
Probes	Yes	Yes	Yes
Irrelevants	Yes	Yes	Yes
Targets	Yes	Yes	*No*
Inducers	*No*	Yes	Yes

Note: Overview of the presence (*Yes*) or absence (*No*) of the item types (probes, irrelevants, targets, and inducers) as described in the text. Note that the procedural differences between task versions concern the target and inducer items exclusively. All three tasks are fully identical in procedure in every other respect.

purposes (task version can be chosen; trial number reduced to five in every phase; no restrictions for error rates).

2.1.1 | Participants

This experiment was run on Figure Eight (www.figure-eight.com; formerly known as CrowdFlower), an online crowdsourcing platform where participants from anywhere in the world can register to complete small online tasks (Peer, Samat, Brandimarte, & Acquisti, 2015). Hence, this website may also be used to offer participation in online experiments by providing a link to the task to be completed (e.g., Kleinberg & Verschuere, 2015). People registered on this site as "contributors" complete many such tasks, and their performance may be rated after the completion of the tasks by the "customers" who offered those tasks. Based on these ratings, contributors are categorized into three levels, where contributors with best ratings are categorized as "Level 3." When creating a new task, a customer (in this case, the current authors) may choose the lowest level of contributors that are allowed to take the task. We set this to "Level 3"; hence, only such "Level 3" contributors were allowed to participate in the study. We opened slots for 310 participants for our experiment, paying 1.20 USD per completed task. The task could only be completed in one uninterrupted time from one IP address: Another attempt from an IP address that was already stored with a completed task resulted in a warning prompt on the first page of the task that did not allow continuation. Possibly due to simultaneous starting times, 313 participants completed the task (see dropout rates in Appendix A).

Each participant was randomly assigned to perform one of the three RT-CIT versions: Target-CIT, E-CIT, or Inducer-CIT. Each participant was also randomly assigned to the *guilty* or *informed innocent* condition. In the guilty condition, the probes were participants' self-reported autobiographical identity details (e.g., their country of origin), simulating a guilty suspect. In the informed innocent condition, the probes were randomly chosen, originally irrelevant identity details, but these participants were informed about these details in advance (simulating an innocent suspect exposed to information leakage; see Section 2.1.2).

Our exclusion criteria were at least 50% accuracy for each of the following item categories: targets, self-referring inducers, and other-referring inducers (Lukács, Kleinberg, & Verschuere, 2017). Furthermore, exclusion criteria were at least 75% overall accuracy for main items (probe or irrelevant items; see preregistration). Only two participants had to be excluded based on these criteria, both only for too low accuracy to self-referring inducers (one from E-CIT and one from Inducer-CIT condition). However, a further 39 participants were excluded due to not recalling correctly the probes at the end of the task (see Section 2.1.2): 18 from Target-CIT (three guilt and 15 informed innocent), 14 from E-CIT (five guilty and nine informed innocent), and seven from Inducer-CIT (three guilty and four informed innocent).[1]

[1]It is perhaps noteworthy that even a few of the guilty participants (11 out of 157), who were supposed to provide their true autobiographical details, could not recall them correctly. Although these persons may have made an incorrect selection by accident, it is also possible that they in fact did not provide their true details honestly, but rather chose them randomly. Hence, it may be advisable to implement this check in all future tasks.

This left 272 participants: 52 in Target-CIT guilty (M_{age} ± SD_{age} = 36.7 ± 11.4; 71.2% male), 36 in Target-CIT informed innocent (M_{age} ± SD_{age} = 37.8 ± 9.9; 61.1% male), 43 in E-CIT guilty (M_{age} ± SD_{age} = 33.0 ± 9.1; 62.8% male), 41 in E-CIT informed innocent (M_{age} ± SD_{age} = 33.8 ± 10.6; 56.1% male), 51 in Inducer-CIT guilty (M_{age} ± SD_{age} = 34.8 ± 8.3; 56.9% male), and 49 in Inducer-CIT informed innocent (M_{age} ± SD_{age} = 33.5 ± 10.7; 65.3% male) groups. Thus, some of the groups have resulted in a lower sample size than our aim of around 50 subjects.

2.1.2 | Procedure

Before beginning the experiment, all participants agreed to the informed consent in order to proceed further. (Both in the informed consent and on the Figure Eight site, the information included the rule, in boldface font, that at least an upper-intermediate English knowledge is required.) Participants then provided demographic information and chose, from dropdown menus, the three autobiographical details: country of origin, date of birth (month and day), and favorite animal. (Countries served as high salient and animals as low salient, see, e.g., Verschuere & Kleinberg, 2015; and birthdays as intermediate in saliency, as found by Lukács & Ansorge, 2019.) This was followed by the very short (3-min) LexTALE English competency test (Lemhöfer & Broersma, 2012), in which 60 words are presented, among which 40 are real English words, whereas 20 are nonwords, and the instruction is to decide, for each word, whether it is an actual English word or not. This test was implemented as described at www.lextale.com, with the only difference that a 4-s time limit applied to each response to curb possible cheating (i.e., looking up the words online or in a dictionary during the task). The LexTALE minimum score for upper intermediate (B2) level is 60% accuracy (Lemhöfer & Broersma, 2012, p. 341). Consequently, those who did not achieve a score above our more lenient threshold of 55% clearly did not have the required English skill and therefore were automatically disqualified and redirected to the Figure Eight website. This screening was important due to the tasks' (presumed) reliance on semantic associations, which requires a clear understanding of basic English (Lukács & Ansorge, 2019), and then followed one of the CITs as described below.

Item selection

Participants were informed that the following task simulates a lie detection scenario, during which they should try to hide their identities. They were also told that they may actually not see their own details in the task, in which case they are in the "innocent" condition, simulating an innocent suspect. They were then presented a short list of randomly chosen items within each of the three categories in the task (countries, dates, and animals). The items on these lists never contained any of the probes (the actual identity details of a given participant), but, within each category, they had the closest possible character length to the given probe (depending on the list of available items), and none of them started with the same letter (except in case of months). In case of countries, if the probe included a space (e.g., "New Zealand" or "Czech Republic"), the items on this list were all

chosen to include a space as well. The participants were asked to choose any (but a maximum of two per category) items that were personally meaningful to them or in any way appeared different from the rest of the items on those lists. Subsequently, the items for the task were randomly selected from the nonchosen items (as this assures that the irrelevants were indeed irrelevant).

For participants in the guilty condition, their self-reported identity details served as the probe, in each of the three categories, whereas the four irrelevants and one target (where this applied) were randomly chosen from among the nonchosen items. (The target was of course not used in the Inducer-CIT; see section.) For a participant in the informed innocent condition, six items were selected for each of the three categories. Out of these six, one was randomly assigned to be a probe, whereas the remaining five served as irrelevants and target (where applicable). Thus, in either condition and in any of the CIT versions, there were altogether three probes and 12 irrelevants, whereas only in the Target-CIT and E-CITs, there were an additional three targets (one per block, see below) as well.

Information leakage

Following the item selection, all innocent participants were informed of the selected probes on a dedicated "background information" page, where they were described a person who "committed a serious crime, but is now hiding his true identity," and they were told that they are one of the suspects (see full text in Appendix B; see also the very similar "background story" in Lukács, Gula, et al., 2017). The country of origin, date of birth, and favorite animal of this person were pointed out repeatedly. On the next page, all participants had to type in all these three details correctly in order to proceed with the test. If any of the entered items was incorrect, the participant received a warning and was redirected to the background information page.

Targets

Next, participants in the Target-CIT and E-CIT versions were presented their three targets and were asked to memorize these items in order to recognize them as requiring a different response during the following task. On the next page, participants were asked to recall the memorized items and could proceed only if they selected these items correctly from a dropdown menu. If any of the entered items was incorrect, the participant received a warning and was redirected to the previous page in order to have another look at the same items. (For the Inducer-CIT, this target learning phase was omitted.)

Task designs

In each RT-CIT task, the items were presented one by one in the center of the screen, and participants had to categorize them by pressing one of two keys ("E" or "I") on their keyboard. The design of the Target-CIT replicated the regular RT-based CIT (mainly Kleinberg & Verschuere, 2015; Seymour, Seifert, Shafto, & Mosmann, 2000; Verschuere & Kleinberg, 2015). Namely, participants were told that pushing the "I" key means "YES," they recognize the item, whereas pushing the "E" key means "NO," they do not recognize the item—and they were correspondingly instructed to say YES to the targets and NO to all other words (i.e., both the irrelevants and the probes). In case of the E-CIT and Inducer-CIT, the description was slightly modified to focus on familiarity: Participants were told that pushing the "I" key means that the displayed item is "FAMILIAR" to them, whereas pushing the "E" key means that the item is "UNFAMILIAR" to them. Participants in these groups also had to categorize nine different inducers: three referring to familiarity ("FAMILIAR," "RECOGNIZED," and "MINE") had to be categorized as familiar ("I" key), whereas the other six referring to unfamiliarity ("UNFAMILIAR," "UNKNOWN," "OTHER," "THEIRS," "THEM," and "FOREIGN") had to be categorized as unfamiliar ("E" key). In case of the E-CIT (but not in the Inducer-CIT), participants were also instructed to respond FAMILIAR to the targets. In both E-CIT and Inducer-CIT, all other words (irrelevants and probes) had to be categorized as unfamiliar.

In previous studies, reminder captions were displayed on the screen throughout the task (e.g., "Recognize?" or "Familiar to you?" on the top of the screen), but, to avoid any related potential confounds, we simply omitted any of these captions altogether. Arguably, this could hardly have any relevant effect, but in order to demonstrate this, we ran a smaller preliminary within-subject experiment (using one guilty E-CIT group only; n = 52), briefly presented in Appendix C, which shows that indeed the presence or absence of captions makes no difference.

The intertrial interval (i.e., between the end of one trial and the beginning of the next) always randomly varied between 100 and 300 ms. In case of a correct response, the next trial followed. In case of an incorrect response or no response within the given time limit, the caption "WRONG" or "TOO SLOW" in red color appeared, respectively, below the stimulus for 400 ms, followed by the next trial.

The main task was preceded by a comprehension check and two practice tasks. The check served to ensure that the participant had fully understood the task. The items consisted of 12–21 randomly ordered trials, including 10–12 different main items (two probes and eight irrelevants and, for the Target-CIT and E-CITs, two targets; each of which was randomly chosen from one out of the three categories, without replacement) and each of the nine possible inducers for the E-CIT and Inducer-CIT version. During the comprehension check, participants had plenty of time (10 s) to choose a response. However, each trial required a correct response. In case of an incorrect response, the participant immediately got a corresponding feedback, was reminded of the instructions, and had to repeat this check. This check guaranteed that the eventual differences (if any) between the responses to the probe and the responses to the irrelevants were not due to misunderstanding of the instructions or any uncertainty about the required responses in the eventual task.

In the following first practice task, the response window was longer than in the main task (2 s instead of 800 ms), whereas the second practice task had the same design as the main task. Both practice tasks consisted of 9–14 trials, in a way that two successive tasks always contained all of the possible items in the task (for Target-CIT: three probes, 12 irrelevants, three targets; for E-CIT: nine inducers

in addition; and also nine additional inducers, but no targets, for Inducer-CIT). In either practice task, in case of too few valid responses, the participants received a corresponding feedback, were reminded of the instructions, and had to repeat the practice task. The requirement was a minimum of 60% valid responses (correct key between 150 and 800 ms) for each of the following item types (when the given type existed in the given CIT version): targets, self-referring inducers, other-referring inducers, and main items (probes and irrelevants together).

Note that in previous online experiments, the exclusion was set to 50%, which is, however, chance level, and seemed too low to us. Also, probe and irrelevants were previously treated separately. In previous experiments, the separate check for probes was included primarily to ensure that the participant had understood the task, but here, this was already fully ensured through the comprehension check.

The main task, in each test, contained three blocks, each for one separate category (countries, dates, or animals; in random order). In each block, each probe, irrelevant, and target (where applicable) were repeated 18 times (hence, altogether 54 probe, 216 irrelevant, and 54 target trials). From the Inducer-CIT, target trials were omitted (leaving 54 probe and 216 irrelevant trials). Within each of the three blocks, the order of the items was randomized in groups: first, all five or six items (one probe, four irrelevants, and, where applicable, one target) in the given category were presented in a random order, and then the same six items were presented in another random order (but with the restriction that the first item in the next group was never the same as the last item in the previous group).

In the E-CIT and Inducer-CIT, inducers were placed among these items in a random order, but with the restrictions that an inducer trial was never followed by another inducer trial, and each of the nine inducers (three self-referring and six other-referring) preceded each of the three probes, three targets (for E-CIT, but not for Inducer-CIT), and 12 irrelevants exactly one time. (Thus, 9 × 18 = 162 inducers were presented in the E-CIT, and 162 out of the 324 other items were preceded by an inducer. Similarly, 9 × 15 = 135 inducers were presented in the Inducer-CIT, and 135 out of the 270 other items were preceded by an inducer.)

Final comprehension check

At the end of the test, to verify the proper understanding of the lie detection scenario simulation and the unimpaired awareness of the correct details up to this point, each participant had to select the probes from the list of all possible items in each category (country, date, and animal). Again, in case of guilty participants, the probes were simply their own details (hence, the instruction read: "please select again the truly self-related details that you yourself gave in the very beginning"). More importantly, in case of informed innocent participants, these were the details they were informed about in the beginning, in the frame of the background information (hence, the instruction read: "please select below the details of the criminal as it was described in the beginning"). All participants who gave any of the details incorrectly were excluded from the analyses (see Section 2.1.1).[2] This is a new (but preregistered) exclusion method, the purpose of which was to ensure that the participants in the informed innocent conditions were indeed properly informed about the probe items. (Nonetheless, for the sake of treating data in all conditions similarly, the same method was applied to the guilty conditions as well.)

After the task, there was a short survey where participants rated the personal importance of the items used in the task (their country of origin, birthday, and favorite animal; on a scale from one to six, where one is "entirely unimportant" and six is "very important"), and finally, the participants were given a brief explanation about the purpose of the study.

2.1.3 | Data analysis

We conducted preregistered analyses, unless explicitly specified otherwise. For examining the main questions, the dependent variable was the probe-minus-irrelevant correct RT mean (correct probe RT mean minus correct irrelevant RT mean, per each participant) used in (a) an analysis of variance (ANOVA) with between-subjects factors Knowledge (guilty or informed innocent) and Version (Target-CIT, E-CIT, or Inducer-CIT),[3] (b) comparisons of areas under the curves (AUCs; see below) between each two of the three task versions (to assess the overall detection accuracy of each version in the information leakage scenario), and (c) *t* tests on the RT mean differences between probes and irrelevants in each of the three informed innocent conditions (to see whether there is a significant effect at all). For secondary analyses, in Appendix D, we report (a) a mixed ANOVA to explore the potential effects of item saliency (countries vs. animals) and its interactions, and (b) all tests described so far (regarding probe and irrelevant RT means) with probe and irrelevant accuracy rates rather than correct RTs.

On the request of reviewers of an earlier version of this manuscript, we (a) added Bayesian analyses to all *F* and *t* tests, and (b) we calculated AUCs for each of the CIT versions with simulated naive, uninformed innocents (contrasted to both guilty and informed innocent participants). For each simulation, we took a randomly generated sample of 100 participants from a normal distribution with a mean of zero (*rnorm* function in R, with *set.seed* at 100). The *SD*s for the Target-CIT and E-CIT were based on the data for the same respective CIT versions in the very similar study of Lukács, Kleinberg, and Verschuere (2017). The results were reanalyzed using the criteria in the present paper and excluding the item category of "favorite color"

[2]All participants who failed on any of the details for the first time were excluded from all analyses. Nonetheless, as an additional check, these participants were once warned about the incorrect selections and were asked to try again. Over half of them (24 out of 39) selected the probes correctly at this second time, indicating that, although they may have been uncertain or confused by this question, they had been properly informed. After this second check, regardless of failed selections, all participants were allowed to finish the task and receive payment.

[3]One reviewer insisted that we run this exploratory ANOVA instead of the following two preregistered tests: (a) a *t* test between the guilty E-CIT and guilty Inducer-CIT version (in order to estimate the difference between their efficiency in case of uninformed innocents) and (b) a one-way analysis of variance (ANOVA) for informed innocent conditions only (to explore possible differences in susceptibility to information leakage). However, which particular method was used here made no difference for the conclusions. The full original, preregistered analysis is uploaded to https://osf.io/zr39m/.

(as the remaining categories of countries, dates, and animals correspond exactly to those in the present paper). The resulting *SD* for the Target-CIT was 16.0 ms and 12.6 ms for the E-CIT. For the Inducer-CIT, we used an average of these two *SD*s: 14.3 ms.[4]

Areas under the curves

To assess the efficiency of discriminating between guilty and informed innocent conditions, we calculated AUCs (a diagnostic efficiency measure, for binary classification, that takes into account the distribution of all predictor values; Rice & Harris, 2005; Zou, O'Malley, & Mauri, 2007) for receiver operating characteristics (ROCs). The AUC can range from 0 to 1, where .5 means chance level classification and 1 means flawless classification (i.e., all guilty and informed innocent classifications can be correctly made based on the given predictor variable, at a given cutoff point).

Effect sizes

In order to demonstrate the magnitude of the observed effects for *F* tests, partial eta-squared (η_p^2) values are shown along with their 90% CIs (Steiger, 2004). We report Welch-corrected *t* tests (Delacre, Lakens, & Leys, 2017) and Cohen's *d* values as standardized mean differences and their 95% CIs (Kelley, 2018; Lakens, 2013). We used the conventional alpha level of .05 for all statistical significance tests.

Bayesian analysis

We report Bayes factors using the default r-scale of 0.707 (Morey & Rouder, 2018). The Bayes factor is a ratio between the likelihood of the data fitting under the null hypothesis and the likelihood of fitting under the alternative hypothesis (Jarosz & Wiley, 2014; Wagenmakers, 2007). For example, a Bayes factor (BF) of 3 means that the obtained data are three times as likely to be observed if the alternative hypothesis is true, whereas a BF of 0.5 means that the obtained data are twice as likely to be observed if the null hypothesis is true. Here, for more readily interpretable numbers, we denote Bayesian factors as BF_{10} for supporting alternative hypothesis and as BF_{01} for supporting null hypothesis. Thus, for example, $BF_{01} = 2$ again means that the obtained data are twice as likely under the null hypotheses than under the alternative hypothesis. Typically, BF = 3 is interpreted as the minimum likelihood ratio for "substantial" evidence for either the null or the alternative hypothesis (Jeffreys, 1961).

For all analyses, RTs below 150 ms were excluded. For RT analyses, only correct responses were used. Accuracy was calculated as the number of correct responses divided by the number of all trials (after the exclusion of those with an RT below 150 ms).

2.2 | Results

2.2.1 | Group-level response time analysis

All means and *SD*s of individual RT means, for the different stimuli types, in all guilty and innocent conditions, are given in Table 2.

[4]Variations of this procedure make very little difference in the obtained AUCs.

We conducted an ANOVA, with between-subjects factors Knowledge (guilty vs. informed innocent) and Version (Target-CIT, E-CIT, and Inducer-CIT), on probe-minus-irrelevant RT mean differences (Table 2; Figure 1). The main effect of Knowledge was significant (larger effect for guilty), $F(1, 266) = 23.72$, $p < .001$, $\eta_p^2 = .082$, 90% CI [.036, .138], $BF_{10} = 533.19$, as well as the main effect of Version, $F(2, 266) = 25.51$, $p < .001$, $\eta_p^2 = .161$, 90% CI [.096, .223], $BF_{10} = 8.75 \times 10^6$. The interaction, however, was not statistically significant, with indeterminate BF, $F(2, 266) = 1.72$, $p = .182$, $\eta_p^2 = .013$, 90% CI [0, .039], $BF_{01} = 1.71$.

Follow-up *t* tests for the Version main effect shows that participants (guilty and informed innocent together) have significantly larger probe-minus-irrelevant RT differences in the E-CIT than in either the Target-CIT, $t(162.5) = 5.27$, $p < .001$, $d_{between} = 0.80$, 95% CI [0.49, 1.11], $BF_{10} = 3.57 \times 10^4$, or in the Inducer-CIT, $t(137.1) = 5.61$, $p < .001$, $d_{between} = 0.83$, 95% CI [0.53, 1.13], $BF_{10} = 4.17 \times 10^5$. However, the difference between the Target-CIT and Inducer-CITs is not significant, $t(162.0) = 0.31$, $p = .759$, $d_{between} = 0.04$, 95% CI [−0.24, 0.33], $BF_{01} = 6.02$.

To test the effect of information leakage on each CIT version separately, we performed paired sample *t* tests between the correct probe RT means and correct irrelevant RT means within each informed innocent condition (for effect sizes, see Table 2). This probe-minus-irrelevant difference proved statistically significant only in case of the E-CIT, $t(40) = 5.66$, $p < .001$, $BF_{10} = 1.20 \times 10^4$. There was no such difference in the Inducer-CIT, with BF substantially supporting null hypothesis, $t(48) = 0.39$, $p = .701$, $BF_{01} = 6.00$, and, despite a small but notable effect size, it was not significant in the Target-CIT, with an indeterminate BF, $t(35) = 1.21$, $p = .234$, $d_{within} = 0.20$, $BF_{01} = 2.85$.

2.2.2 | Individual classification

Probe-minus-irrelevant differences in RT mean were used as predictor variables to calculate AUCs, which are shown for each condition in Table 2. Diagnostic accuracy was very modest for both the Target-CIT and the E-CIT, but notably better for the Inducer-CIT. Using DeLong's test for the statistical comparison of two AUC values (DeLong, DeLong, & Clarke-Pearson, 1988; Robin et al., 2011), we found that the AUC for the Inducer-CIT was significantly higher than that of the Target-CIT protocol, $D(157.94) = 2.43$, $p = .016$, as well as higher than that of the E-CIT, $D(156.79) = 2.14$, $p = .034$. There was no difference between the AUCs of the E-CIT or Target-CIT, $D(169.88) = 0.32$, $p = .751$.

The simulated AUCs reflect what was expected based on the comparisons of probe-minus-irrelevant RT means. For guilty versus simulated uninformed innocent participants, they were .652, 95% CI [.561, .744] for Target-CIT, .885, 95% CI [.825, .945] for E-CIT, and 0.778, 95% CI [.705, .851] for Inducer-CIT. For informed innocent versus simulated uninformed innocent participants, they were .533, 95% CI [.418, .648] for Target-CIT, .761, 95% CI [.662, .861] for E-CIT and .525, 95% CI [.429, .621] for Inducer-CIT.

TABLE 2 RT means and related Cohen's *d*s and areas under the curves in Experiment 1

	SP Target-CIT		E-CIT		Inducer-CIT	
	Guilty	Informed innocent	Guilty	Informed innocent	Guilty	Informed innocent
Probe	442 ± 43	440 ± 47	511 ± 40	510 ± 40	447 ± 46	435 ± 44
Irrelevant	433 ± 41	436 ± 42	484 ± 41	493 ± 33	433 ± 42	435 ± 45
Target	519 ± 40	517 ± 36	562 ± 35	570 ± 36	–	–
Self-referring	—	—	591 ± 32	599 ± 38	582 ± 31	582 ± 31
Other-referring	—	—	530 ± 45	540 ± 39	510 ± 52	520 ± 49
P–I	8.8 ± 15.5	4.2 ± 20.6	28.0 ± 21.6	17.1 ± 19.4	14.4 ± 12.4	0.6 ± 11.1
d_{within}	0.57 [0.27, 0.86]	0.20 [−0.13, 0.53]	1.30 [0.89, 1.70]	0.88 [0.52, 1.24]	1.16 [0.80, 1.51]	0.06 [−0.23, 0.34]
$d_{between}$	0.25 [−0.18, 0.68]		0.53 [0.09, 0.96]		1.17 [0.74, 1.59]	
AUC	.609 [.484, .734]		.637 [.518, .756]		.798 [.711, .884]	

Note: Means and *SD*s (in the format of M ± *SD*) for individual RT means for *Probe* (item presumed to be the participant's own detail), *Irrelevant* (other details in the categories as the probe), *Target* (that designated irrelevant details that require different response), *Self-referring* (self-referring inducers), *Other-referring* (other-referring inducers), and *P* – *I* (individual probe minus irrelevant values). Dashes indicate inapplicable cases: no inducers in the Target-CIT and no targets in the Inducer-CIT. Cohen's *d* effect sizes (with 95% CIs in brackets): d_{within} for probe-minus-irrelevant differences and $d_{between}$ for differences between guilty and informed innocent for each CIT version. *AUC*: Area under the curve (i.e., classification accuracy between the guilty and informed innocent participants of each CIT version).

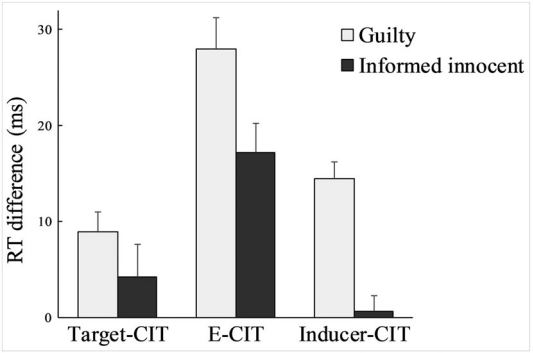

FIGURE 1 Means and *SE*s of individual probe-minus-irrelevant response time (RT) mean differences in Experiment 1 for the guilty participants (with their own details as probes) and for the informed innocent participants (with random details as probe, but informed about it), in the three CIT versions: Target-CIT (with probes, irrelevants, and targets), E-CIT (with probes, irrelevants, targets, and self-referring inducers), and Inducer-CIT (with probes, irrelevants, and other-referring inducers)

2.3 | Discussion

In Experiment 1, we have shown that the RT-CIT is vulnerable to information leakage when the targets were present in the E-CIT, whereas it was not affected when using the Inducer-CIT with no targets but only inducers. It was also shown that both the Target-CIT and E-CIT provide very modest classification accuracies in such a case, whereas the Inducer-CIT version remains fairly efficient.

However, unfortunately, the relatively small difference between probe and irrelevant RTs did not reach statistical significance in case of the informed innocents with the Target-CIT version, so the effect of information leakage was not yet directly shown in this version. The small effect size (and the consequent *p* value) is arguably due to the suboptimal nature of this version, which shows much smaller effect sizes than the other versions even in case of guilty participants. As it can be observed in Figure 1 or Table 2, it appears to be proportionally just as much affected by information leakage as the E-CIT version. Nonetheless, to show unambiguous proof of the effect of information leakage on the Target-CIT, we conducted a second experiment using a multiple-probe (MP) protocol Target-CIT.

In Experiment 1, we used the single-probe (SP) protocol Target-CIT, where each category is presented in separate blocks (see right panel in Figure 2). The SP has several practical advantages, such as applicability even in case of a limited number of probe items (Podlesny, 2003), compatibility with common test procedures and scoring algorithms (Krapohl, 2011), and sequential testing to narrow down possibilities (Lukács, Kleinberg, & Verschuere, 2017). Consequently, practitioners currently also consider the SP protocol to be the only viable option (Ogawa, Matsuda, Tsuneoka, & Verschuere, 2015). Furthermore, in our experiment, we use SP protocol for the E-CIT and Inducer-CIT versions, which would be unnecessarily complex with a corresponding MP protocol (Lukács, Kleinberg, & Verschuere, 2017). Hence, for comparability, we wanted to use the SP protocol in the Target-CIT as well.

However, in each block, the inducers constitute additional nine items, among which there are three that have to be categorized as "targets" (i.e., opposite to the probe and irrelevants). Therefore, in respect of the number of different items, the MP protocol with its multiple items in each block (in particular, three targets) is in fact more comparable with the Inducer-CIT. Furthermore, it has been repeatedly shown that the MP protocol clearly outperforms the SP protocol in the (RT-based) Target-CIT (Eom, Sohn, Park, Eum, & Sohn, 2016; Lukács, Kleinberg, & Verschuere, 2017; Verschuere & Kleinberg, 2015). Therefore, from the practical perspective, the future use of the suboptimal SP Target-CIT seems unlikely. Finally, we expect the

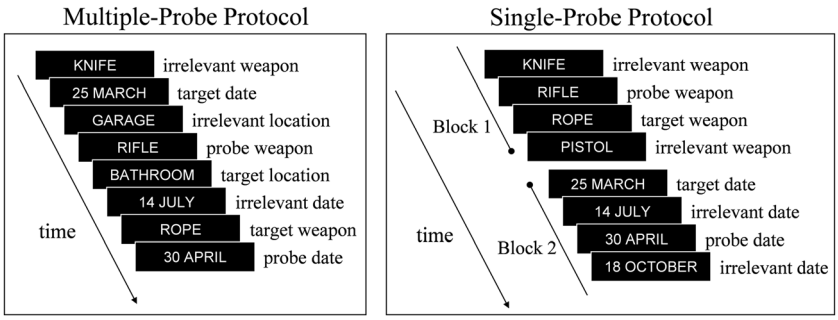

FIGURE 2 The multiple-probe and the single-probe protocols of the CIT (Lukács, Kleinberg, & Verschuere, 2017) illustrated with a hypothetical murder case. (For brevity, the larger ratio of irrelevants is not proportionately represented in this illustration)

larger effect to also be present in case of informed innocents, leading to increased statistical power.

Consequently, in Experiment 2, we ran the same study as in Experiment 1, but with the MP Target-CIT. In the MP protocol (see left panel in Figure 2), items related to the different categories (e.g., weapons, locations, and dates) are completely intermixed within blocks. Additionally, we again included the Inducer-CIT to ensure that the null finding among informed innocents from Experiment 1 is replicable.

3 | EXPERIMENT 2

3.1 | Methods

Experiment 2 was preregistered at https://osf.io/ys4a2, with all behavioral data (original and aggregated) available via https://osf.io/zr39m/, along with the entire original task.

3.1.1 | Participants

Participants were sampled via Figure Eight as in Experiment 1, with the same procedure and exclusion criteria. However, in this experiment, to ensure that we have the intended minimum sample size in each condition, we preregistered a procedure to sample more participants (n + 5 more for n missing) in case of less than 50 valid completions (after all exclusions, see below) in any given condition. Following that procedure, we first opened 220 slots. Then we opened six more slots for the guilty MP Target-CIT (as it initially had 49 valid completions), six for the guilty Inducer-CIT versions (also 49 initial valid completions), 17 more for the informed innocent MP Target-CIT (38 initial valid completions), and, finally, again for the informed innocent MP Target-CIT, seven more (as it still only had 48 valid completions).

Altogether, 257 participants completed the task (see dropout rates in Appendix A). Each participant was randomly assigned to perform one of the two RT-CIT versions: MP Target-CIT, or Inducer-CIT. Each participant was also randomly assigned to the guilty or informed innocent condition. Three participants were excluded for too low accuracy to targets (one from guilty, two from informed innocent Target-CIT), and four for too low accuracy to self-referring inducers (three from guilty and one from informed innocent Inducer-CIT). Further 37 participants were excluded due to not recalling correctly the probes at the end of the task: 28 from MP Target-CIT (three guilty and 25 informed innocent) and nine from Inducer-CIT (six guilty and three informed innocent).

This left 212 participants: 56 in Target-CIT guilty (M_{age} ± SD_{age} = 37.2 ± 10.2; 64.3% male), 52 in Target-CIT informed innocent (M_{age} ± SD_{age} = 38.4 ± 9.6; 61.5% male), 53 in Inducer-CIT guilty (M_{age} ± SD_{age} = 34.8 ± 11.6; 69.8% male), and 51 in Inducer-CIT informed innocent (M_{age} ± SD_{age} = 33.5 ± 8.8; 60.8% male) groups.

3.1.2 | Procedure

The procedure corresponded exactly to that in Experiment 1, except a very slight modification in the description of required responses. Namely, in Experiment 1, the instruction text related the response keys to recognition in case of Target-CIT and to familiarity in case of Inducer-CIT (see Section 2.1.2). To make the instructions more straightforward, in Experiment 2, we removed these explanatory references and simply instructed participants to press the given keys when the corresponding items were displayed (e.g., press the key "I" for the target details and press "E" for everything else, without further explanation about what these keypresses mean). Furthermore, there were no reminder captions displayed at any point (i.e., not even during the first two practice phases as in Experiment 1).

The arrangement of items was also identical to that in Experiment 1, except that in the MP Target-CIT, instead of one category per block, each block contained an equal number of items intermixed from each category (countries, dates, or animals). Within each of the three blocks, the order of the items was randomized in groups: First, all 18 items (three probes, 12 irrelevants, and four targets) were presented in a random order (but with the restriction that target trials were never followed by another target trial, and probe trials were never followed by another probe trial). Then the same 18 items were presented in another random order (but with the restriction that the first item in the next group was never the same as the last item in the previous group).

3.1.3 | Data analysis

We again conducted preregistered analyses, unless explicitly specified otherwise. We used the probe-minus-irrelevant correct RT mean as dependent variable in (a) an ANOVA with between-subjects factors

Knowledge (guilty vs. informed innocent) and Version (MP Target-CIT vs. Inducer-CIT) and (b) a comparison of the AUCs between the two task versions with a one-sided DeLong's test, expecting higher AUC for the Inducer-CIT. In case of this test, we did not preregister expected direction: However, this direction is clear from the context and also based on the results of Experiment 1. Finally, also not preregistered, we exploratorily repeated the same DeLong's test using the Inducer-CIT data from both experiments to increase statistical power.

To directly test the effect of information leakage in each of the two CIT versions separately, we performed one-sided *t* tests for informed innocents only, expecting slower probe responses. To test the null hypothesis, we conducted corresponding one-sided Bayesian *t* tests as well. Because we expect this probe-minus-irrelevant effect in case of informed innocents to be proportional to the effect in case of guilty participants, we adjusted the r-scale based on the probe-irrelevant Cohen's *d* obtained from the corresponding guilty condition: The r-scale (for informed innocent) is calculated as half the value of the Cohen's *d* (from guilty)—see the preregistration for the precise implementation in R. However, for completeness, we also report BF values with default r-scale (which differ only very slightly).

As supplementary analyses in Appendix D, we again report (a) a mixed ANOVA to explore the potential effects of item saliency and its interactions and (b) all tests described so far with probe and irrelevant accuracy rates rather than correct RTs.

Same as in Experiment 1, on the request of reviewers, (a) we added Bayesian analyses to all *F* and *t* tests (with the default r-scale of 0.707), and (b) we calculated AUCs for the MP Target-CIT versions with simulated uninformed innocents. The simulation was performed in the same manner, with *SD* based on the results for the same version in the study of Lukács, Kleinberg, and Verschuere (2017), from which we obtained an *SD* of 17.2 for the MP Target-CIT.

3.2 | Results

3.2.1 | Group-level response time analysis

All means and *SD*s of individual RT means, for the different stimuli types, in all guilty and informed innocent conditions, are given in Table 3.

We conducted an ANOVA, with between-subjects factors Knowledge (guilty vs. informed innocent) and Version (MP Target-CIT vs. Inducer-CIT), on probe-minus-irrelevant RT mean differences. We found significant main effects for both Knowledge (larger values for guilty), $F(1, 208) = 19.43$, $p < .001$, $\eta_p^2 = .085$, 90% CI [.034, .150], $BF_{10} = 846.24$, and Version (larger values for MP Target-CIT), $F(1, 208) = 9.24$, $p = .003$, $\eta_p^2 = .043$, 90% CI [.009, .095], $BF_{10} = 8.57$. There was, however, no significant Knowledge × Version interaction, $F(1, 208) = 0.30$, $p = .582$, $\eta_p^2 = .001$, 90% CI [0, .022], $BF_{01} = 6.28$ (see Figure 3).

The one-sided paired sample *t* tests between the probe RT means and irrelevant RT means within each informed innocent condition (for effect sizes, see Table 3) showed a significant effect in case of MP Target-CIT, $t(51) = 3.97$, $p < .001$, $BF_{10} = 1.53 \times 10^8$ (default r-scale of 0.707), adjusted $BF_{10} = 1.50 \times 10^8$ (r-scale of 0.380), but no significant effect for Inducer-CIT, with BFs again supporting null hypothesis, $t(50) = -0.02$, $p = .507$, $BF_{01} = 5.73$ (default r-scale of 0.707), adjusted $BF_{01} = 5.87$ (r-scale of 0.353).

3.2.2 | Individual classification

Probe-minus-irrelevant differences in RT mean were used as predictor variables to calculate AUCs, which are shown for each condition in Table 3. Diagnostic accuracy was very modest for the MP Target-CIT (almost as low as for the SP Target-CIT in Experiment 1), but notably better for the Inducer-CIT. However, the comparison of two

TABLE 3 RT means and related Cohen's *d*s and areas under the curves in Experiment 2

	MP Target-CIT			Inducer-CIT
	Guilty	Informed innocent	Guilty	Informed innocent
Probe	509 ± 44	500 ± 39	447 ± 50	437 ± 45
Irrelevant	491 ± 35	491 ± 40	435 ± 47	437 ± 44
Target	583 ± 36	581 ± 33	—	—
Self-referring	—	—	576 ± 35	578 ± 36
Other-referring	—	—	507 ± 53	521 ± 53
P–I	18.7 ± 24.5	9.0 ± 16.3	12.4 ± 17.6	0.0 ± 11.1
d_{within}	0.76 [0.46, 1.06]	0.55 [0.26, 0.84]	0.71 [0.40, 1.00]	0.00 [−0.28, 0.27]
$d_{between}$	0.47 [0.08, 0.85]			0.85 [0.45, 1.25]
AUC	.633 [.527, .739]			.734 [.637, .831]

Note: Means and *SD*s (in the format of M ± *SD*) for individual RT means for *Probe* (item presumed to be the participant's own detail), *Irrelevant* (other details in the categories as the probe), *Target* (that designated irrelevant details that require different response), *Self-referring* (self-referring inducers), *Other-referring* (other-referring inducers), and *P–I* (individual probe minus irrelevant values). Dashes indicate inapplicable cases: no inducers in the Target-CIT and no targets in the Inducer-CIT. Cohen's *d* effect sizes (with 95% CIs in brackets): d_{within} for probe-minus-irrelevant differences and $d_{between}$ for differences between guilty and informed innocent for each CIT version. *AUC*: Area under the curve (i.e., classification accuracy between the guilty and informed innocent participants of each CIT version).

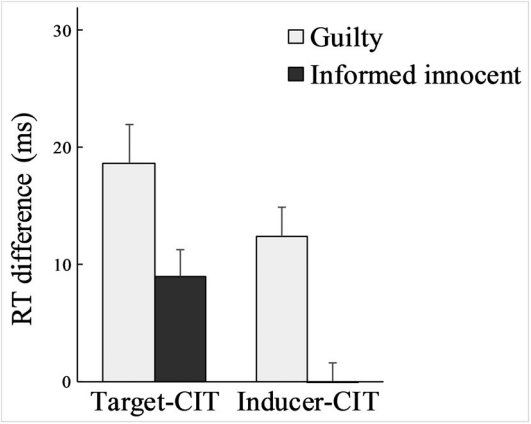

FIGURE 3 Means and *SE*s of individual probe-minus-irrelevant response time (RT) mean differences in Experiment 1 for the guilty participants (with their own details as probes) and for the informed innocent participants (with random details as probe, but informed about it), in the two CIT versions: Target-CIT (with probes, irrelevants, and targets), Inducer-CIT (with probes, irrelevants, and other-referring inducers)

AUC values using a one-sided DeLong's test showed fell short of statistical significance, $D(208.86) = 1.38$, $p = .085$. Finally, for increased statistical power, we repeated this test using all Inducer-CIT data from both experiments (combined AUC: .765 [.700, .830]), in which case the Inducer-CIT did clearly prove to have higher AUC than the MP Target-CIT, $D(188.12) = 2.07$, $p = .020$.

For guilty versus simulated uninformed innocent participants, the AUCs were .740, 95% CI [.654, .826] for MP Target-CIT and .706, 95% CI [.621, .791] for Inducer-CIT. For informed innocent versus simulated uninformed innocent participants, they were .662, 95% CI [.571, .754] for MP Target-CIT and .493, 95% CI [.399, .587] for Inducer-CIT.

3.3 | Discussion

In this second experiment, we successfully replicated the finding that the Inducer-CIT is resistant to information leakage. Furthermore, as it was clearly shown in case of the E-CIT, but only indicative in case of the SP Target-CIT, we have now shown in the MP Target-CIT that the presence of the target leads to a significant probe-minus-irrelevant difference in case of informed innocents.

We found no significant interaction between the factors Knowledge (guilty vs. informed innocent) and CIT Version (MP Target-CIT vs. Inducer-CIT) when using RT means. This is because the probe-minus-irrelevant differences for guilty participants in case of the MP Target-CIT were large enough to still create a substantial difference between guilty and informed innocent participants, despite of the effect of information leakage. This difference was similar to that in case of the Inducer-CIT, which was not susceptible to information leakage, but nonetheless had comparatively low probe-minus-irrelevant differences for guilty participants.

We did, however, find a significant interaction between these same factors when using accuracy rates (see Appendix D). In fact, the effect of information leakage on accuracy rate probe-minus-irrelevant differences in the MP Target-CIT was so large that informed innocents had significantly larger such differences than guilty participants. This may also indicate a speed–accuracy tradeoff (Heitz, 2014). Namely, informed innocent participants may have focused on giving fast responses to the probe, instead of accurate responses—hence, the effect of information leakage is observable primarily in accuracy rates and less in mean correct RTs.

Furthermore, even with the RT measure, despite no significant interaction of group *means*, we found that the Inducer-CIT had a higher AUC (i.e., it could better discriminate guilty from informed innocent than the MP Target-CIT): This is because, in addition to the in fact slightly larger differences between the group *means* of guilty and informed innocent individual probe-minus-irrelevant differences (though not statistically significant), the Inducer-CIT also exhibited smaller *variance* of these individual probe-minus-irrelevant differences, in both guilty and informed innocent groups (see *SD*s in Table 3). That is, for both groups (guilty and informed innocent), the predictor variables (individual probe-minus-irrelevant differences) were more narrowly distributed and, hence, allowed less overlap between guilty and informed innocent participants. The higher variance in the MP Target-CIT may be due to the presence of the targets' different influences on different persons. In particular, the subjective meaning of the targets to any given examinee may modulate their effects (Suchotzki, De Houwer, Kleinberg, & Verschuere, 2018).

In any case, we have unambiguously proven our two main points: (a) the significant adverse effect of information leakage on the standard CIT (when target is present) and (b) the absence of this effect when the target is not present in the CIT task.

4 | GENERAL DISCUSSION

In the present paper, we have shown that the RT-CIT is sensitive to information leakage and, therefore, cannot be effectively applied in field settings where the probe may be known to innocent suspects. We have also shown a possible remedy in form of an RT-CIT relying primarily on associations. At the same time, our findings also provide insight into the mechanism of the CIT, supporting previously proposed theories (Lukács, Gula, et al., 2017; Lukács, Kleinberg, & Verschuere, 2017).

We did so by comparisons between three essential designs: (a) the Target-CIT (both SP and MP), where the target items foster response conflict for any relevant rare item among the nontargets; (b) the Inducer-CIT, where items referring to the self-relatedness of the probe induce semantic item-category associations, causing response conflict for self-related probes due to their semantic incongruence with the other-referring inducers and irrelevants; and (c) E-CIT, where both target and inducer items were included, and, hence, the two mechanisms exerted a combined influence. As expected, these versions were differentially affected by the two conditions, guilty (where

the probes were self-related and closely familiar to the participants) and informed innocent (where probes were not self-related, but recognized as relevant to the task). Although the probe recognition increased probe-minus-irrelevant differences of informed innocents in the Target-CIT and E-CIT versions in the presence of targets, there was no such difference in the Inducer-CIT version.

The classification accuracies of both the Target-CIT (AUC = .61 for SP in Experiment 1 and .63 for MP in Experiment 2) and E-CIT (AUC = .64) fall short of practical value in the deception detection context (National Research Council, 2003). The Inducer-CIT version, however, provides an appreciable accuracy (AUC = .80 in Experiment 1 and .73 in Experiment 2), although, unfortunately, it also proved less effective than the E-CIT version (see our simulated AUC of .89 for E-CIT or the actual AUC of .94 in the very similar study of Lukács, Kleinberg, & Verschuere, 2017). Moreover, it is noteworthy that the Inducer-CIT achieved, despite the information leakage, a substantially higher classification accuracy than the SP Target-CIT in comparable studies even without leakage (e.g., Lukács & Ansorge, 2019; Lukács, Kleinberg, & Verschuere, 2017; Verschuere & Kleinberg, 2015).

In conclusion, this new version stands as a viable future option in case of informed innocent participants, regardless of whether the information leakage is a known fact or a mere suspected possibility. Naturally, further research into the topic will be needed, as well as independent replications. Given the multiple facets of this method (several item types, each with its own parameters of categorization, visual display, timing, semantic attributes, etc.), there are abundant opportunities for improvements as well. For example—as for an aspect relevant to the rest of this discussion as well—the semantic dimension of the inducers did not relate directly to the probes. In particular, the self-referring fillers were "familiar," "recognized," and "mine"—out of which "mine" refers most closely to the self-related probes (as opposed to irrelevants), whereas, on the other hand, "recognized" may be perceived as referring to items recognized as relevant, with an unnecessary focus on the recognition factor. An improvement, therefore, could consist of using exclusively self-related and/or directly probe-related concepts (e.g., "home country," "birthday," "my favorite," and "my own"). For theoretical purposes, this may be contrasted in an experiment with more strictly recognition-focused inducers ("recognized," "relevant," etc.).

This connects to another important point, namely, the role of the target item. It was shown, in case of informed innocents, that the inclusion of the target in the E-CIT elicits a large probe-minus-irrelevant effect in sharp contrast to the otherwise identical Inducer-CIT. As described in detail in Section 1, we attribute this to the fact that the probe's uniqueness and relevance is shared by the target. However, one may also consider the target as an item that controls for semantic category: In the Inducer-CIT, participants may recognize the item category before fully processing the relevance of the specific item (in particular, the probe) and, therefore, could categorize it based partly on its category (e.g., always press left key for countries, regardless of whether they are irrelevants or probes), diminishing probe-minus-irrelevant differences to some extent. This could explain the smaller probe-minus-irrelevant differences for Inducer-CIT compared with E-CIT in Experiment 1. This adverse factor in the Inducer-CIT for guilty cases cannot be circumvented, because it is simultaneously the essence of protection against the adverse effect of information leakage in innocent cases.

However, the targets may have an additional role as visual control items as well: In the Inducer-CIT, probe and irrelevant items may be partly discriminated based on visual cues, such as character length or spaces. Consequently, a potential improvement would be to present the inducers in a format more visually similar to that of the current probe and irrelevant items: For example, complementing them with filler characters for corresponding length, inserting spaces, or, in case of dates, appending random numbers to them (e.g., "MINE 19" or "OTHER 07"). In essence, all measures that prevent a fast classification of inducers on the basis of their visual features alone should in turn foster their semantic processing and the resulting conflict between inducers and probes among guilty participants.

As described in the previous paragraphs and as evident from the differences between the E-CIT and Inducer-CIT (Table 2 and Figure 1), it is understood that the targets contribute to the probe-minus-irrelevant effect in case of guilty participants along with the use of inducers. Nonetheless, our findings also show that, in turn, the inducers also facilitate the probe-minus-irrelevant difference when combined with the use of targets: In case of informed innocents (where only recognition of relevance plays a part), inducers alone created no probe-minus-irrelevant difference (d_{within} = 0.06 in the Inducer-CIT), whereas targets alone created only small to medium effects (d_{within} = 0.20 in the SP Target-CIT, 0.55 in the MP Target-CIT), but with the use of both, there is a large effect (d_{within} = 0.88 in the E-CIT). We attribute this to the increased cognitive load elicited by the involvement of two semantic dimensions if not even tasks (i.e., the discrimination between self- vs. other-relatedness and the discrimination between response meanings of targets and irrelevants as separate instances of items not related to the self) inviting closer attention and, thereby, deeper processing of the stimuli (as argued in more detail, but not tested, by Lukács, Kleinberg, & Verschuere, 2017).

Thus, we have provided evidence for the impact of targets on probe-minus-irrelevant differences and for the influence of semantically associated inducers on the same differences, as two separate components that can work independently—but that can also interact, if combined, producing an effect that surpasses the effect of either component alone, for both of the examined scenarios (guilty and informed innocent cases).

Because it limits the applicability of the method (Podlesny, 2003), in this paper, we have so far only presented this probe-minus-irrelevant effect as unfavorable in case of innocent examinees. However, importantly, there are cases where this may in fact be desirable: For example, a witness (either a bystander or an accomplice) may falsely deny the recognition of a suspect. In that case, the guiltiness of the examinee is not in question, but merely the fact of recognition. The present study demonstrates that the E-CIT is likely to be effective in this situation as well—although such specific scenarios will require further investigations.

4.1 | Limitations

Same as in almost all deception research experiments, "guilt" and "innocence" were merely simulated in our study. Although the personal relevance of the presented autobiographical details arguably resembles the relevance of real-life incriminating items, the extent of applicability is yet to be explored. On the one hand, in a specific situation very similar to the one simulated in the present study, authorities may test the true identity of the person, in which case the results may be assumed comparable with those in our study (regarding higher stakes at hand, see Kleinberg & Verschuere, 2016). On the other hand, the relevance of crime-related items (such as a murder weapon), which may be contributed to by the various emotions related to the actually committed crime (guilt, suspense, etc.), would be very difficult to simulate in a controlled experiment and may require field studies in the future.

Relatedly, in our study, autobiographical identity details were the objects of the test, with inducers referring to familiarity and ownership. On the one hand, the same principle may be adapted to other scenarios: for example, in case of a murderer's gun ("my gun") or for a stolen object ("my loot"). On the other hand, this may not always be as straightforward as in case of identity details: For example, a thief might not consider a stolen object as his own property. This could be examined in the future, along with potential use of action related expressions as inducers, such as "I stole" and "they stole," etc. (Lukács, Gula, et al., 2017).

Our study did not assess how well people can remember leaked information. Our study concerned the consequences of remembered leaked details. Therefore, we simply excluded informed innocent participants who did not correctly recall the probe items about which they were informed. Because in real-world situations innocents would probably forget some of the leaked information (and because such cases were excluded in the present study), we tested the impact of leaked information under relatively conservative conditions.

All in all, future research should explore the applicability of our findings in differing scenarios, including possibly more realistic simulations of information leakage. This may also include testing the method's potential susceptibility to countermeasures (Hu, Rosenfeld, & Bodenhausen, 2012; Verschuere, Prati, & De Houwer, 2009).

The online setting of our experiment may allow more noise in the data than strictly controlled lab experiments due to (a) potentially suboptimal computer hardware used by participants, (b) varying environmental factors (e.g., different lighting conditions or disruptive noises), and (c) participants' lower motivation to pay close attention and perform the task properly. However, there have been numerous studies that explored the reliability of online psychological research (e.g., Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014), in particular the validity of online RT research (e.g., Germine et al., 2012; Hilbig, 2016; McGraw, Tew, & Williams, 2000; for the HTML5/JavaScript framework as in the present study, see Reimers & Stewart, 2015) and recently even specifically the validity of the RT-CIT in online settings (Kleinberg & Verschuere, 2015; Verschuere & Kleinberg, 2015)— all of which unanimously conclude that online RT research such as in our study is a sound alternative to conventional lab studies and that results obtained in this environment closely reflect those obtained in strictly controlled laboratory conditions. Note also that, in our experiment, the key dependent variable (probe-minus-irrelevant differences) was obtained, for each participant, via a within-subject comparison, which also serves as a control for external influences. Finally, online research has its advantages as well: in particular, a highly diverse international sample that provides a broad demonstration of generalizability and also more closely reflects the test results of possible criminal suspects than a study involving only university students as in typical lab studies.

ACKNOWLEDGMENTS

We thank Anna Walker and Matthew Pelowski for proofreading. Gáspár Lukács is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute for Basic Psychological Research at the University of Vienna. This funding source had no role or involvement related to the present study.

AUTHORS' CONTRIBUTIONS

Concept, study design, and data acquisition by G. L.; manuscript drafted by G. L., revised by U. A.

CONFLICT OF INTEREST

The authors have no financial or personal conflicts of interest.

ORCID

Gáspár Lukács https://orcid.org/0000-0001-9401-4830

REFERENCES

Agosta, S., & Sartori, G. (2013). The autobiographical IAT: A review. *Frontiers in Psychology*, *4*, 519. https://doi.org/10.3389/fpsyg.2013.00519

Azizian, A., Freitas, A. L., Watson, T. D., & Squires, N. K. (2006). Electrophysiological correlates of categorization: P300 amplitude as index of target similarity. *Biological Psychology*, *71*(3), 278–288. https://doi.org/10.1016/j.biopsycho.2005.05.002

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, *38*(6), 977–997. https://doi.org/10.1002/ejsp.487

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, *134*(4), 477–492. https://doi.org/10.1037/0033-2909.134.4.477

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.011

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. *International Review of Social Psychology*, *30*(1), 92. https://doi.org/10.5334/irsp.82

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837–845. https://doi.org/10.2307/2531595

Eom, J.-S., Sohn, S., Park, K., Eum, Y.-J., & Sohn, J.-H. (2016). Effects of varying numbers of probes on RT-based CIT accuracy. *International Journal of Multimedia and Ubiquitous Engineering*, *11*(2), 229–238. https://doi.org/10.14257/ijmue.2016.11.2.23

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, *105*(2), 203. https://doi.org/10.5195/JMLA.2017.88

Gamer, M., & Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology*, *47*(2), 355–364. https://doi.org/10.1111/j.1469-8986.2009.00933.x

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659. https://doi.org/10.1037/a0023589

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*(150), 1–19. https://doi.org/10.3389/fnins.2014.00150

Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hu, X., Rosenfeld, J. P., & Bodenhausen, G. V. (2012). Combating automatic autobiographical associations: The effect of instruction and training in strategically concealing information in the Autobiographical Implicit Association Test. *Psychological Science*, *23*(10), 1079–1085. https://doi.org/10.1177/0956797612443834

Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2015). Basic level category structure emerges gradually across human ventral visual cortex. *Journal of Cognitive Neuroscience*, *27*(7), 1427–1446. https://doi.org/10.1162/jocn_a_00790

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1), 2–9. https://doi.org/10.7771/1932-6246.1167

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, England: Clarendon Press.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16–32. https://doi.org/10.1037/0022-3514.91.1.16

Kelley, K. (2018). MBESS: The MBESS R package. R package version 4.4.3. Retrieved from https://CRAN.R-project.org/package=MBESS

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

Kleinberg, B., & Verschuere, B. (2016). The role of motivation to avoid detection in reaction time-based concealed information detection. *Journal of Applied Research in Memory and Cognition*, *5*(1), 43–51. https://doi.org/10.1016/j.jarmac.2015.11.004

Krapohl, D. J. (2011). Limitations of the Concealed Information Test in criminal cases. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 151–170). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.009

Kraut, R. (1980). Humans as lie detectors. *Journal of Communication*, *30*(4), 209–218. https://doi.org/10.1111/j.1460-2466.1980.tb02030.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–243. https://doi.org/10.3758/s13428-011-0146-0

Lukács, G. (2019). CITapp—a response time-based Concealed Information Test lie detector web application. *Journal of Open Source Software*, *4*(34), 1179. https://doi.org/10.21105/joss.01179

Lukács, G., & Ansorge, U. (2019). Methodological improvements of the Association-Based Concealed Information Test. *Acta Psychologica*, *194*, 7–16. https://doi.org/10.1016/j.actpsy.2019.01.010

Lukács, G., Grządziel, A., Kempkes, M., & Ansorge, U. (2019). Item roles explored in a modified P300-based CTP Concealed Information Test. *Applied Psychophysiology and Biofeedback*. https://doi.org/10.1007/s10484-019-09430-6

Lukács, G., Gula, B., Szegedi-Hallgató, E., & Csifcsák, G. (2017). Association-based Concealed Information Test: A novel reaction time-based deception detection method. *Journal of Applied Research in Memory and Cognition*, *6*(3), 283–294. https://doi.org/10.1016/j.jarmac.2017.06.001

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., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, *110*, 56–65. https://doi.org/10.1016/j.ijpsycho.2016.10.010

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

Marchand, Y., Inglis-Assaff, P. C., & Lefebvre, C. D. (2013). Impact of stimulus similarity between the probe and the irrelevant items during a card-playing deception detection task: The "irrelevants" are not irrelevant. *Journal of Clinical and Experimental Neuropsychology*, *35*(7), 686–701. https://doi.org/10.1080/13803395.2013.819837

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science*, *11*(6), 502–506. https://doi.org/10.1111/1467-9280.00296

Meijer, E. H., klein Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta-analysis of skin conductance, respiration, heart rate, and P300 data: CIT meta-analysis of SCR, respiration, HR, and P300. *Psychophysiology*, *51*(9), 879–904. https://doi.org/10.1111/psyp.12239

Meijer, E. H., Smulders, F. T. Y., & Wolf, A. (2009). The contribution of mere recognition to the P300 effect in a Concealed Information Test. *Applied Psychophysiology and Biofeedback*, *34*(3), 221–226. https://doi.org/10.1007/s10484-009-9099-9

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.12-4.2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

National Research Council (2003). *Polygraph and lie detection*. Washington, D.C., US: The National Academies Press. Retrieved from: http://www.nap.edu/openbook.php?record_id=10420

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York, NY, US: Psychology Press.

Ogawa, T., Matsuda, I., Tsuneoka, M., & Verschuere, B. (2015). The Concealed Information Test in the laboratory versus Japanese field practice: Bridging the scientist-practitioner gap. *Archives of Forensic Psychology*, *1*(2), 16–27.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Peer, E., Samat, S., Brandimarte, L., & Acquisti, A. (2015). Beyond the Turk: An empirical comparison of alternative platforms for online behavioral research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2594183

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*(3). Retrieved from: https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2003/podlesny.htm

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, *47*(2), 309–327. https://doi.org/10.3758/s13428-014-0471-1

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Seymour, T. L., & Schumacher, E. H. (2009). Electromyographic evidence for response conflict in the exclude recognition task. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 71–82. https://doi.org/10.3758/CABN.9.1.71

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

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*(2), 164–182. https://doi.org/10.1037/1082-989X.9.2.164

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

Varga, M., Visu-Petra, G., Miclea, M., & Buş, I. (2014). The RT-based Concealed Information Test: An overview of current research and future perspectives. *Procedia - Social and Behavioral Sciences*, *127*, 681–685. https://doi.org/10.1016/j.sbspro.2014.03.335

Verschuere, B., & Kleinberg, B. (2015). ID-check: Online Concealed Information Test reveals true identity. *Journal of Forensic Sciences*, *61*(S1), S237–S240. 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., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking in the autobiographical implicit association test. *Psychological Science*, *20*(4), 410–413. https://doi.org/10.1111/j.1467-9280.2009.02308.x

Verschuere, B., Suchotzki, K., & Debey, E. (2015). Detecting deception through reaction times. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception detection: Current challenges and new approaches* (pp. 269–291). Oxford, UK: John Wiley & Sons.

Visu-Petra, G., Varga, M., Miclea, M., & Visu-Petra, L. (2013). When interference helps: Increasing executive load to facilitate deception detection in the Concealed Information Test. *Frontiers in Psychology*, *4*, 146. https://doi.org/10.3389/fpsyg.2013.00146

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, *115*, 654–657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929

How to cite this article: Lukács G, Ansorge U. Information leakage in the Response Time-Based Concealed Information Test. *Appl Cognit Psychol*. 2019;33:1178–1196. https://doi.org/10.1002/acp.3565

APPENDIX A

Dropout rates

In the first E-CIT article (Lukács, Kleinberg, & Verschuere, 2017), due to technical reasons, dropout rates were reported based only on an estimate: the number of participants who successfully completed the first practice phase. Here, we are more precise, reporting the number of all participants who began the test (following the English test and before starting the first practice, where conditions began to differ; Table A1).

APPENDIX B

The full text for the background information was (with example probes *Sweden*, *December 22*, and *otter*, which were always inserted automatically):

TABLE A1 Dropout rates

	Guilty			Informed innocent		
	Target-CIT	E-CIT	Inducer-CIT	Target-CIT	E-CIT	Inducer-CIT
Experiment 1						
Initial	68	120	110	62	136	89
Dropout	13 (19%)	72 (60%)	56 (51%)	11 (18%)	86 (63%)	36 (40%)
Experiment 2						
Initial	74	—	123	116	—	104
Dropout	15 (20%)	—	65 (53%)	39 (34%)	—	51 (49%)

Note: Initial numbers of participants who reached at least the page of the task instructions (where the conditions begin to differ) and the number of those who dropped out during the test (i.e., did not submit a completed test; also, in parentheses, dropout as the percentage of the initial number).

> *There is a person from Sweden, who committed a serious crime, but is now hiding his true identity. We do not know how this person looks, but we know that he or she was born on December 22, and his or her favorite animal is the otter.*
> *We have several suspects, including you.*
> *Now we will do a lie detection test to see whether or not you are from Sweden, born on December 22, or have otter as your favorite animal.*

And on the next page:

> *Importantly, regardless of whether or not these details (Sweden, December 22, otter) are yours, you do not want us to think that you are guilty, and therefore you should deny that these details are yours. In the task, you should treat these details just the same as any other - so that you prove that you are innocent.*
> *However, make sure you do not forget the criminal's details! You will again be asked about them at the end of the task.*

APPENDIX C

Caption display in the Response Time-Based Concealed Information Test

In the original study introducing the E-CIT (Lukács, Kleinberg, & Verschuere, 2017), the differences in caption display during the test were noted as a possible minor confound. Namely, the Target-CIT had, as reminders, the following captions displayed throughout the task: "Recognize?" at the top of the screen, "YES = e" on the left side, "NO = i" on the right side (where "e" and "i" refer to the corresponding response keys on a standard keyboard; same as in previous experiments). The E-CIT had slightly modified captions to correspond to the use and the concept of the familiarity-related inducers: "Familiar to you?" at the top, "FAMILIAR = e" on the left, "UNFAMILIAR = i" on the right. Nonetheless, in a supplementary experiment, it was shown that the results do not change even when using the same captions (always familiarity-related) in all versions.

Here, we add that, arguably, having or not having these captions hardly makes any difference in the first place, and to preclude even the suspicion of any related confound, we shall simply omit them altogether. Still, to consider any potential effects, one could say that, although in the beginning the captions may facilitate the understanding of the task, eventually these additional items on the screen may just cause distraction from the critical stimuli presented in the center. Therefore, before we proceeded with the main objectives of our study (information leakage), we ran a smaller experiment testing any potential difference between captions-on and captions-off versions, within-subject, using only one group: guilty participants in the E-CIT, which is the most complex version, and, therefore, assumed to be most susceptible to be influenced by either distractions or facilitated comprehension.

The data were collected the exact same way as in the main experiment, except that there was only one condition (E-CIT guilty) and four blocks: either starting with two blocks with captions displayed on the screen, or with two blocks with no captions displayed, and the last two blocks with the opposite display type (i.e., no captions, or with captions, respectively). Each two blocks contained once countries, once animals, in this fixed blocked order. This task, too, is available via https://osf.io/zr39m/, including a demonstration (task version can be chosen; trial number reduced to five in every phase; no restrictions for IP or for error rates). Out of the 57 participants who completed the task (whereas 94 dropped out), five had to be excluded: one for too low accuracy to targets, one for too low accuracy to familiarity-referring inducers, and three for not recalling correctly the probe items at the end of the task. This left 52 participants, out of which 25 started with captions off (M_{age} ± SD_{age} = 34.6 ± 9.6; 72.0% male) and 27 with captions on (M_{age} ± SD_{age} = 34.2 ± 11.6; 63.0% male).

Paired sample *t* tests showed no difference between the captions-on and captions-off displays, neither in RT means, $t(51) = 0.77$, $p = .446$, $d_{within} = 0.11$, 95% CI [−0.17, 0.38], $BF_{01} = 5.01$, nor in accuracy rates, $t(51) = 0.72$, $p = .477$, $d_{within} = 0.10$, 95% CI [−0.17, 0.37], $BF_{01} = 5.19$; see Table B1. The order (starting with captions on or with captions off) had no effect on either measure ($p > .4$).

We can conclude that no differences were found between the captions-on and captions-off versions for the means of the probe-

TABLE B1 RT means, accuracy rates, and related Cohen's *d*s

	Captions on	Captions off
Means (ms)		
Probe	504 ± 48	506 ± 39
Irrelevant	478 ± 45	478 ± 37
Target	557 ± 40	558 ± 43
Self-referring	588 ± 34	591 ± 37
Other-referring	523 ± 50	526 ± 40
P–I	26.1 ± 23.8	28.0 ± 19.4
d_{within}	1.09 [0.75, 1.44]	1.45 [1.05, 1.83]
Accuracies (%)		
Probe	96.4 ± 4.6	96.7 ± 4.3
Irrelevant	98.8 ± 1.5	98.6 ± 1.6
Target	81.2 ± 9.6	81.8 ± 11.5
Self-referring	76.6 ± 11.1	77.8 ± 12.2
Other-referring	95.1 ± 3.8	94.7 ± 4.2
P–I	−2.42 ± 4.12	−1.92 ± 3.84
d_{within}	−0.59 [−0.88, −0.29]	−0.50 [−0.79, −0.21]

Note: Means and *SD*s (in the format of M ± *SD*) for individual RT means for *Probe* (item presumed to be the participant's own detail), *Irrelevant* (other details in the categories as the probe), *Target* (that designated irrelevant details that require different response), *Self-referring* (self-referring inducers), *Other-referring* (other-referring inducers), and *P–I* (individual probe minus irrelevant values). Cohen's d_{within} for probe-minus-irrelevant differences.

minus-irrelevant (RT or accuracy) differences. Nonetheless, the captions-off version's larger probe-minus-irrelevant effect size (Table B1) may still let us assume that the captions do cause some minor distraction, and, therefore, if any preference is to be taken, the captions-off version should be favored.

APPENDIX D

Supplementary analyses

Experiment 1

Saliency

We examined the effect of saliency and its possible interactions across the CIT versions for probe-minus-irrelevant RT means. In a three-way ANOVA, with Saliency (high-salient countries vs. low-salient animals) as within-subject factor and Version (Target-CIT, E-CIT, Inducer-CIT) and Knowledge (guilty vs. informed innocent) as between-subjects factors, neither the three-way interaction, $F(2, 266) = 2.96$, $p = .053$, $\eta_p^2 = .022$, 90% CI [0, .054], $BF_{10} = 1.31$, nor the Saliency × Version interaction, $F(2, 266) = 1.17$, $p = .312$, $\eta_p^2 = .009$, 90% CI [0, .031], $BF_{01} = 10.85$, was significant. The Saliency main effect, however, was significant in the expected direction, with probe-minus-irrelevant differences larger for high-salient (country) items than for low-salient (animal) items, $F(1, 266) = 13.13$, $p < .001$, $\eta_p^2 = .047$, 90% CI [.014, .094], $BF_{10} = 29.74$. Furthermore, there was a significant Saliency × Knowledge interaction: The observed Saliency effect was only present in the guilty conditions and not in the informed innocent condition, $F(1, 266) = 38.33$, $p < .001$, $\eta_p^2 = .126$, 90% CI [.070, .188], $BF_{10} = 10.50 \times 10^6$; see Figure D1. This is unsurprising: The saliency refers to the personal importance of the probe (which is higher in case of countries of origins than in case of favorite animals), but, in case of participants only *informed* of the probe, the probes in any item categories have equal importance in that they are recognized as relevant to the test (but have no further personal importance).

Accuracy rates

We report statistical results for accuracy rates in the same manner as for RTs. All means and *SD*s of individual accuracy rates, for the different stimuli types, in all guilty and informed innocent conditions, are given in Table D1.

We conducted an ANOVA, with between-subjects factors Knowledge (guilty vs. informed innocent) and Version (Target-CIT, E-CIT, and Inducer-CIT), on probe-minus-irrelevant accuracy rate differences (Table D1). The main effect of Version was significant (larger effect for guilty), $F(2, 266) = 12.37$, $p < .001$, $\eta_p^2 = .085$, 90% CI [.036, .138], $BF_{10} = 2.33 \times 10^3$. However, neither the effect of Knowledge was significant, $F(1, 266) = 0.18$, $p = .672$, $\eta_p^2 = .001$, 90% CI [0, .015], $BF_{01} = 6.27$, nor the interaction, $F(2,266) = 1.69$, $p = .186$, $\eta_p^2 = .013$, 90% CI [0, .038], $BF_{01} = 3.78$. Follow-up *t* tests for the Version main effect show that participants (guilty and informed innocent together) have significantly larger probe-minus-irrelevant accuracy rate differences in the E-CIT than in either the Target-CIT, $t(106.3) = 3.91$, $p < .001$, $d_{between} = 0.60$, 95% CI [0.29, 0.90], $BF_{10} = 204.96$, or in the Inducer-CIT, $t(105.1) = 3.29$, $p = .001$, $d_{between} = 0.49$, 95% CI [0.19, 0.78], $BF_{10} = 43.53$. However, the difference between the Target-CIT and Inducer-CITs is not significant, $t(184.3) = 1.29$, $p = .200$, $d_{between} = 0.19$, 95% CI [−0.10, 0.47], $BF_{01} = 2.93$.

Probe-minus-irrelevant differences in accuracy rates were used as predictor variables to calculate AUCs for each condition (Table D1).

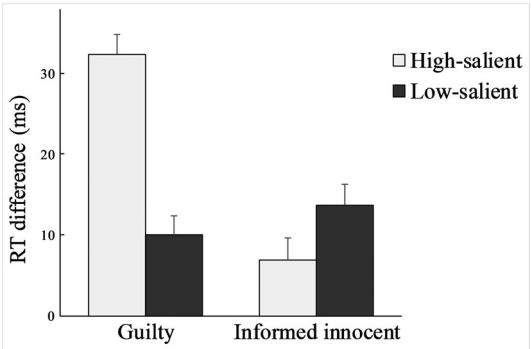

FIGURE D1 Means and *SE*s of individual probe-minus-irrelevant reaction time (RT) mean differences (i.e., correct probe RT means minus correct irrelevant RT means) in Experiment 1. *High-salient*: item category in which the probe is highly personally important. *Low-salient*: item category in which the probe is less personally important. *Guilty:* participants with their own details as probes. *Innocent*: participants with random details as probe, but informed about it. (In this figure, all three task versions are merged together)

TABLE D1 Accuracy rates and related Cohen's *d*s and areas under the curves in Experiment 1

	Target-CIT CIT			E-CIT		Inducer-CIT
	Guilty	Informed innocent	Guilty	Informed innocent	Guilty	Informed innocent
Probe	98.9 ± 1.9	98.2 ± 2.7	96.5 ± 3.5	95.5 ± 5.8	98.4 ± 2.5	99.2 ± 1.2
Irrelevant	99.0 ± 1.1	98.7 ± 1.3	98.6 ± 1.5	98.6 ± 1.5	99.4 ± 1.0	99.4 ± 0.8
Target	87.1 ± 9.5	88.0 ± 8.6	79.3 ± 11.0	81.5 ± 8.1	–	–
Self-referring	–	–	74.4 ± 12.2	77.9 ± 10.2	74.8 ± 10.1	77.3 ± 11.6
Other-referring	–	–	94.2 ± 3.9	94.7 ± 4.3	95.0 ± 3.6	96.0 ± 3.4
P–I	−0.08 ± 1.69	−0.51 ± 2.27	−2.07 ± 3.98	−3.05 ± 5.95	−0.99 ± 2.43	−0.25 ± 1.37
d_{within}	−0.05 [−0.32, 0.22]	−0.22 [−0.55, 0.11]	−0.52 [−0.84, −0.20]	−0.51 [−0.84, −0.18]	−0.41 [−0.69, −0.12]	−0.19 [−0.47, 0.10]
$d_{between}$	0.21 [−0.22, 0.63]			0.19 [−0.24, 0.62]		−0.37 [−0.77, 0.02]
AUC	.441 [.319, .563]			.482 [.355, .609]		.554 [.442, .666]

Note: Means and *SD*s (in the format of M ± *SD*) for individual accuracy rates (percentages of correct responses) for *Probe* (item presumed to be the participant's own detail), *Irrelevant* (other details in the same categories as the probe), *Target* (the designated irrelevant details that require a different response), *Self-referring* (self-referring inducers), *Other-referring* (other-referring inducers), *P–I* (individual probe minus irrelevant values). Dashes indicate inapplicable cases: no inducers in the Target-CIT, and no targets in the Inducer-CIT. Cohen's *d* effect sizes (with 95% CIs in brackets): d_{within} for probe-minus-irrelevant differences, $d_{between}$ for differences between guilty and informed innocent for each CIT version. *AUC*: Area under the curve (i.e., classification accuracy between the guilty and informed innocent participants of each CIT version).

There were no significant differences between the accuracy-based AUCs of any two of the three task versions ($p > .4$ for all comparisons).

To test the effect of information leakage on each CIT version separately, we performed paired sample *t* tests between the probe accuracy rates and irrelevant accuracy rates within each informed innocent condition (for corresponding effect sizes, see Table D1). Same as for RT means, this probe-minus-irrelevant difference proved statistically significant only in case of the E-CIT, $t(40) = 3.29$, $p = .002$, $BF_{10} = 15.68$. There was no such significant difference in the Target-CIT, $t(35) = 1.34$, $p = .189$, $BF_{01} = 2.91$, nor in the Inducer-CIT, $t(48) = 1.30$, $p = .199$, $BF_{01} = 3.67$.

Probe-minus-irrelevant differences in accuracy rates were used as predictor variables to calculate AUCs, which are shown for each condition in Table D1. We found no significant difference between any two of the three AUCs ($p > .18$ for all DeLong's test).

Finally, we examined the effects of salience and its possible interactions across the CIT versions for probe-minus-irrelevant accuracy rate differences in a three way ANOVA, with Saliency (high-salient countries vs. low-salient animals) as within-subject factor and Version (Target-CIT, E-CIT, Inducer-CIT) and Knowledge (guilty vs. informed innocent) as between-subjects factors. Neither the main effect of Saliency nor any of its interactions were significant ($p > .15$ for all tests).

Experiment 2

Saliency

We again examined the effect of saliency and its possible interactions across the CIT versions for probe-minus-irrelevant RT means. In a three-way ANOVA, with Saliency (high-salient countries vs. low-salient animals) as within-subject factor and Version (Target-CIT, Inducer-CIT) and Knowledge (guilty vs. informed innocent) as between-subjects factors, the three-way interaction was not significant, $F(1, 208) = 3.79$, $p = .053$, $\eta_p^2 = .018$, 90% CI [.000, .058], $BF_{01} = 1.02$. The Saliency main effect, however, was significant in the expected direction, with probe-minus-irrelevant differences larger for high-salient (country) items than for low-salient (animal) items, $F(1, 208) = 9.71$, $p = .002$, $\eta_p^2 = .045$, 90% CI [.010, .098], $BF_{10} = 10.51$. The Saliency × Version interaction was also significant, though, barely, and with an indeterminate BF, $F(1, 208) = 4.06$, $p = .045$, $\eta_p^2 = .019$, 90% CI [0, .060], $BF_{01} = 1.29$, indicating a somewhat larger effect of Saliency (i.e., larger difference between high- and low-salient categories) in case of the Inducer-CIT. Furthermore, same as in Experiment 1, there was a significant Saliency × Knowledge interaction: The observed Saliency effect was only present in the guilty conditions and not in the informed innocent condition, $F(1, 208) = 16.24$, $p < .001$, $\eta_p^2 = .072$, 90% CI [.026, .134], $BF_{10} = 398.15$; see Figure D2.

Accuracy rates

All means and *SD*s of individual accuracy rates are given in Table D2.

We conducted an ANOVA with between-subjects factors Knowledge (guilty vs. informed innocent) and Version (MP Target-CIT vs.

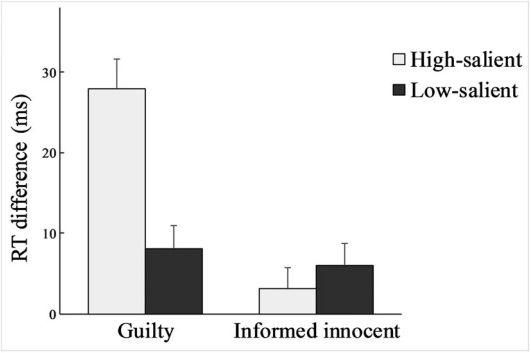

FIGURE D2 Means and *SE*s of individual probe-minus-irrelevant RT mean differences (i.e., correct probe RT means minus correct irrelevant RT means) in Experiment 2. *High-salient*: item category in which the probe is highly personally important. *Low-salient*: item category in which the probe is less personally important. *Guilty:* participants with their own details as probes. *Innocent*: participants with random details as probe, but informed about it. (In this figure, the two task versions are merged together)

Inducer-CIT), on probe-minus-irrelevant accuracy rate differences. We found a significant main effect of Version (larger *negative* probe-minus-irrelevant difference for MP Target-CIT), $F(1, 208) = 15.88$, $p < .001$, $\eta_p^2 = .071$, 90% CI [.025, .132], $BF_{10} = 150.38$. There was no significant Knowledge main effect, $F(1, 208) = 2.63$, $p = .106$, $\eta_p^2 = .012$, 90% CI [0, .048], $BF_{01} = 2.30$. There was, however, a significant Knowledge × Version interaction, $F(1, 208) = 8.83$, $p = .003$, $\eta_p^2 = .041$, 90% CI [.008, .093], $BF_{10} = 5.38$. This interaction indicates that in case of the MP Target-CIT, the negative probe-minus-irrelevant accuracy rate difference was not larger for guilty than for informed innocents, as it normally happens—whereas it was so, as expected, in case of Inducer-CIT (see Table D2). Follow-up *t* tests indicate that both these differences were significant (although, unlike for the interaction, with indeterminate BFs): larger differences in case of informed innocent MP Target-CIT than guilty MP Target-CIT, $t(90.2) = 2.41$, $p = .018$, $BF_{10} = 2.80$, but larger differences in case of guilty Inducer-CIT than informed innocent Inducer-CIT, $t(86.3) = -2.19$, $p = .031$, $BF_{10} = 1.67$ (for effect sizes, see Table D2). This means that, surprisingly, probe-minus-irrelevant accuracy rate differences would better predict *informed* innocence reversely with the MP Target-CIT as compared to what is usually expected: In the MP Target-CIT, informed innocent participants give more incorrect probe responses (as compared with irrelevant responses) than guilty participants.

The one-sided paired sample *t* tests between the probe accuracy rates and irrelevant accuracy rates within each informed innocent condition (for effect sizes, see Table D2) showed a significant effect for both MP Target-CIT, $t(51) = -4.94$, $p < .001$, $BF_{10} = 1.62 \times 10^{12}$ (default r-scale of 0.707), adjusted $BF_{10} = 9.13 \times 10^{11}$ (r-scale of 0.203), and Inducer-CIT, $t(50) = -2.08$, $p = .022$, $BF_{10} = 64.02$ (default r-scale of 0.707), adjusted $BF_{10} = 127.99$ (r-scale of 0.265).

Finally, probe-minus-irrelevant differences in accuracy rates were used as predictor variables to calculate AUCs, which are shown for each condition in Table D2. The AUC for Inducer-CIT was shown significantly higher than that of the Target-CIT, using a one-sided DeLong's test, $D(209.06) = 2.58$, $p = .005$.

TABLE D2 Accuracy rates and related Cohen's *d*s and areas under the curves in Experiment 2

	MP Target-CIT		Inducer-CIT	
	Guilty	Informed innocent	Guilty	Informed innocent
Probe	96.5 ± 4.5	94.2 ± 6.5	98.4 ± 2.0	99.2 ± 1.3
Irrelevant	98.1 ± 1.5	98.2 ± 1.8	99.5 ± 0.8	99.5 ± 0.7
Target	79.6 ± 9.8	78.3 ± 9.3	—	—
Self-referring	—	—	76.7 ± 10.5	79.2 ± 10.1
Other-referring	—	—	96.0 ± 3.8	95.3 ± 4.1
P–I	−1.64 ± 4.04	−3.98 ± 5.80	−1.08 ± 2.04	−0.36 ± 1.24
d_{within}	−0.41 [−0.68, −0.13]	−0.68 [−0.98, −0.38]	−0.53 [−0.82, −0.24]	−0.29 [−0.57, −0.01]
$d_{between}$	0.46 [0.08, 0.84]		−0.43 [−0.82, −0.04]	
AUC	.364 [.258, .470]		.636 [.530, .742]	

Note: Means and *SD*s (in the format of M ± *SD*) for individual accuracy rates (percentages of correct responses) for *Probe* (item presumed to be the participant's own detail), *Irrelevant* (other details in the same categories as the probe), *Target* (the designated irrelevant details that require a different response), *Self-referring* (self-referring inducers), *Other-referring* (other-referring inducers), and *P–I* (individual probe minus irrelevant values). Dashes indicate inapplicable cases: no inducers in the Target-CIT and no targets in the Inducer-CIT. Cohen's *d* effect sizes (with 95% CIs in brackets): d_{within} for probe-minus-irrelevant differences and $d_{between}$ for differences between guilty and informed innocent for each CIT version. *AUC*: Area under the curve (i.e., classification accuracy between the guilty and informed innocent participants of each CIT version).