Association-based Concealed Information Test: A Novel Reaction Time-Based Deception Detection Method

Gáspár Lukács[a,b], Bartosz Gula[b,c], Emese Szegedi-Hallgató[a], Gábor Csifcsák[a,d]

[a] Department of Cognitive and Neuropsychology, Institute of Psychology, Faculty of Arts, University of Szeged, Egyetem u. 2, 6722 Szeged, Hungary

[b] Department of Psychology, University of Klagenfurt, Universitätsstr. 65-67, 9020 Klagenfurt, Austria

[c] Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

[d] Department of Psychology, University of Tromsø, Huginbakken 32, 9037 Tromsø, Norway

Abstract

In recent years, numerous studies were published on the reaction time (RT)-based Concealed Information Test (CIT). However, an important limitation of the CIT is the reliance on the recognition of the probe item, and therefore the limited applicability when an innocent person is aware of this item. In the present paper, we introduce an RT-based CIT that is based on item-category associations: the Association-based Concealed Information Test (A-CIT). Using the participants' given names as probe items and self-referring "inducer" items (e.g., "MINE" or "ME") that establish an association between ownership and responses choices, in Experiment 1 (within-subject design; n = 27), this method differentiated with high accuracy between guilty and innocent conditions. Experiment 2 (n = 25) replicated Experiment 1, except that the participants were informed of the probe item in the innocent condition – nonetheless, the accuracy rate remained high. Implications and future possibilities are discussed.

Key words: Memory Detection, Deception, Concealed Information Test, Reaction Time, Association, Recognition

Technological deception detection methods are widely needed, because without such aid, it is extremely difficult – if not impossible – to tell whether a person is telling the truth or not (Bond & DePaulo, 2006, 2008; Hartwig & Bond, 2011; Kraut, 1980). One frequently researched method is the Concealed Information Test (CIT; Lykken, 1959; Verschuere & Meijer, 2014). The CIT allows to disclose whether an examinee recognizes certain relevant items such as a weapon used in a recent robbery among a set of other objects when he/she actually tries to conceal any knowledge about the criminal case. The recognition of a relevant item can be detected by various means, for instance from increased stress reactions as measured with a polygraph, or, from relatively slower responding to relevant items as assessed with a reaction time-based CIT (RT-CIT). However, the applicability of this test is limited in real life settings, since it cannot be used when an innocent person would also recognize the incriminating item, for example due to information leakage and the consequential increased familiarity with the critical item (Bradley, Barefoot, & Arsenault, 2011). In the present paper, we introduce the Association-based Concealed Information Test (A-CIT), a new RT-based paradigm that aims at identifying concealed knowledge linked to words (e.g., nouns or verbs associated with the crime) just like the RT-CIT (Seymour, Seifert, Shafto, & Mosmann, 2000). However, rather than relying on the recognition of unique items, the A-CIT is based on item-category associations and shares many common features with the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). Before we describe the new method in detail, we shortly present the two approaches that inspired the A-CIT.

The RT-CIT consist of a fast, two-alternative forced choice task, where participants classify the presented stimuli as targets or non-targets by pressing one of two keys. Several, e.g., 6-7 items are presented, among which one is the *probe* item (the item that the guilty person would recognize, e.g., the murder weapon) and the rest are *irrelevant* items (items that are similar to the probe – and thus indistinguishable from the probe for an innocent person). These items are repeatedly shown in a random sequence, and all of them have to be responded to with the same response keys, except the one *target* (irrelevant) item – a randomly selected irrelevant item that has to be answered with the other response key (serving as an 'oddball' in this task). In case of guilty examinees, the answer to the probe will be generally slower (and somewhat more often incorrect) in comparison to the irrelevant items because by recognizing the probe as personally relevant, it will become unique (another 'oddball') and in this respect, more similar to the rarely occurring target item (Varga, Visu-Petra, Miclea, & Buş, 2014; Verschuere & Meijer, 2014; Verschuere, Suchotzki, & Debey, 2015).

The main advantages of the RT-CIT are its low costs and its easy implementation: it can be run using any regular personal computer and takes little time (10-15 minutes). Since the method does not require special equipment, it can very easily be standardized in order to run it in the same manner on any computer, including an immediate automatic analysis of the results (see Verschuere & Kleinberg, 2015).

However, a major limitation of the CIT in connection with any measure (RT, polygraph, EEG, fMRI) is that it uses the recognition of the concealed information as the evidence to classify someone as guilty or not. This makes the test unviable, if the suspect has a way to know the information, i.e., the probe, e.g., in the case of leaked crime details (Bradley et al., 2011; Verschuere & Meijer, 2014). Unfortunately, in the majority of real life scenarios, the probe is indeed known to the suspects – which is the primary reason for the very limited actual field application of the CIT (Ben-Shakhar, 2012; Podlesny, 2003).

The IAT, on the other hand, is not based on recognition, but on item-category associations. There has been a series of studies with IAT-based lie detection, using the IAT basically in its standard format (autobiographical IAT, or aIAT; review: Agosta & Sartori, 2013). As critical items presented during the task, the aIAT uses sentences that each refer to one of two opposing claims about a past event, e.g., having or not having used cocaine (Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008, p. 774). In addition, there are "inducer" items presented in every second trial (i.e., one after each critical item), which are either clearly true or clearly false statements, e.g., "I'm in front of a computer" (true), or "I'm at the beach" (false). Throughout the task, each item has to be responded to with one of two keys on a keyboard, based on the meaning of the item: e.g., having used cocaine with the "e" key, and not having used with the "i" key, while clearly true statements with the "e," and clearly false statements with the "i" key. Due to the strong association between the true critical item and the category of clearly true events, responses are generally faster when the these sentences require the same key press, and slower when the sentences related to true critical events require the same key press as clearly false statements (Sartori et al., 2008; Agosta & Sartori, 2013; Greenwald et al.,1998; Lane, Banaji, Nosek, & Greenwald, 2007). This provides a lie detection method that is highly adaptable to many scenarios, including those where possibly innocent suspects are also aware of all the critical details of a crime, because it is not the recognition of a relevant item that matters, but the association between the critical items and inducers with similarly true or false contents. The studies on the aIAT from the original author show very high accuracy (Agosta & Sartori, 2013), but the accuracies found by independent replications studies are generally lower (see Verschuere, Suchotzki, et al., 2015).

Introducing the Association-based Concealed Information Test

The A-CIT shares similarities with the RT-CIT in that (1) it is designed to detect concealed information, (2) uses simple words as stimuli, and (3) focuses on reaction time differences between probe and irrelevant stimuli. On the other hand, its design, which we briefly introduce below and describe in detail in the Methods, is much more similar to the IAT.

In the A-CIT, there are two kinds of stimuli that appear intermixed in a random order within the same experimental block: first, the critical items (in our experiment, personal names) which includes a probe (the participant’s own name) and several irrelevant items (other personal names), and second, inducer items (expressions describing self-reference or ownership) that are intended to be categorized as phrases that belong to the examinee. The inducers have an important role as they establish an association between certain concepts (here: ownership) and key responses. Participants are asked to make conscious categorization of all stimuli by pressing one of the two response buttons: one explicitly linked to the category in which all of the critical items would truly belong in case of an innocent examinee (“other name” category), while the other related to the inducers(describing self-reference, and belonging to the “my name” category). However, for guilty participants, the probe item is associated with the category of the inducers (here: because the probe is the participant’s name), and thereby this protocol is expected to be suitable for uncovering concealed information (i.e., association between the examinee and the critical item) by yielding altered behavioral measures (accuracy, reaction times) for probe items only.

Experiment 1

The first experiment was run to establish whether our A-CIT can work with an acceptable level of accuracy. Therefore, same as in the case of a regular CIT study, participants in the innocent condition were not aware of which of the given names are the probes (i.e., the event of leaked crime details was not simulated).

Methods

Participants

Pilot testing with earlier versions of the paradigm was performed at the Department of Psychology, University of Szeged, Hungary. The final version of the A-CIT was first tested in Experiment 1, with the voluntary participation of twenty-eight bachelor students enrolled at the Department of Psychology, University of Klagenfurt, Austria (to receive "experiment participation hours" for curriculum requirements). Data from one of these participants was excluded from all analysis due to high error rates in the task (response accuracy over 1.5 interquartile outside the interquartile range), leaving 27 participants (age = 23.22±4.09 years, in the format of MEAN±SD, as also in the rest of this paper; 9 male). The experiment was run with a within-subject design: 14 participants were randomly assigned to first perform the A-CIT in guilty condition, and then the A-CIT in innocent condition, while 13 were assigned to perform the two tasks in the reverse order. The study conformed with the Declaration of Helsinki and was approved by the Institutional Review Board of Department of Psychology, University of Szeged, Hungary.

The Association-based Concealed Information Test Design

In our study, the critical items were five given names (including the participant's own name in the guilty condition). The inducer items were four different expressions referring to own name (e.g., "my name" or "mine").[1] Throughout the task, all these items had to be categorized under two labels: "my name" or "other name." Inducer expressions referring to own name had to be categorized as "my name," while all actual given names had to be categorized as "other name" – since, according to the deception scenario that is simulated in the experiment, the examinee denies that any of the names are his/her own (including the probe, i.e., the one name that we presume to be the examinee's actual name).

[1]All the original German expressions can be found at https://osf.io/k47cg/ in Appendix A, along with their English translation.

Categorization happened through pressing one of two keys, one on the left (“e”), and one on the right (“i”), in accordance with the labels (“my name” and “other name”) that were displayed on the upper part of the screen, one on the left, one on the right. Thus, for example, when an expression referring to the participant’s own name appeared, and the label on the right was “my name,” then the key on the right was to be pressed.

The factually correct category, and therefore the natural association for an irrelevant name is “other name,” while the factually correct category, and therefore the natural association for the person’s own name is “my name.” Consequently, our hypothesis was that due to the conflict between natural associations and task requirements, a guilty person will categorize his/her own name less easily as “other name” as compared to irrelevant names. Thus, since the task always requires each name to be categorized as “other name,” we expected that a guilty person’s responses to his/her own name (i.e., the probe) would be slower, and more often incorrect, than those to the irrelevant items – while in case of an innocent person (whose name does not appear in the test), no substantial differences would be found between the presumed probe and the irrelevant items. This would allow to efficiently distinguish between a guilty and an innocent participant, based on RT and accuracy differences. Furthermore, since this difference is based on item-category association, and not on recognition (such as in the RT-CIT), we would expect that it would not be substantially diminished even in case the probe is known to the examinee.

However, when always pressing the same key for the same category (e.g., if the “my name” label were always in the right corner), the categorization could become automatic: examinees would simply recognize the given names as ones that have to be categorized to one side (e.g., always with the key on the left), regardless whether the name was their own or not, i.e., disregarding the inducer items. To ensure that the meaning of the sides is thoroughly attended throughout the whole task, labels switched or did not switch places at random on each new trial during the task (Meissner & Rothermund, 2013; Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009) – see Figures 1 and 2. Thus, on each trial, participants first had to take a look at the position of the labels and consider their meaning – for example, with “other name” label on the left, and “my name” label on the right, participants had to quickly consider that, on the given trial, items belonging to the “other name” category have to be categorized with the left key, while those belonging to the “my name” category have to be categorized with the right key. This prevented, or at least limited, automatic responding – which could otherwise diminish the differences between the responses to the participant’s own name and the responses to other names.

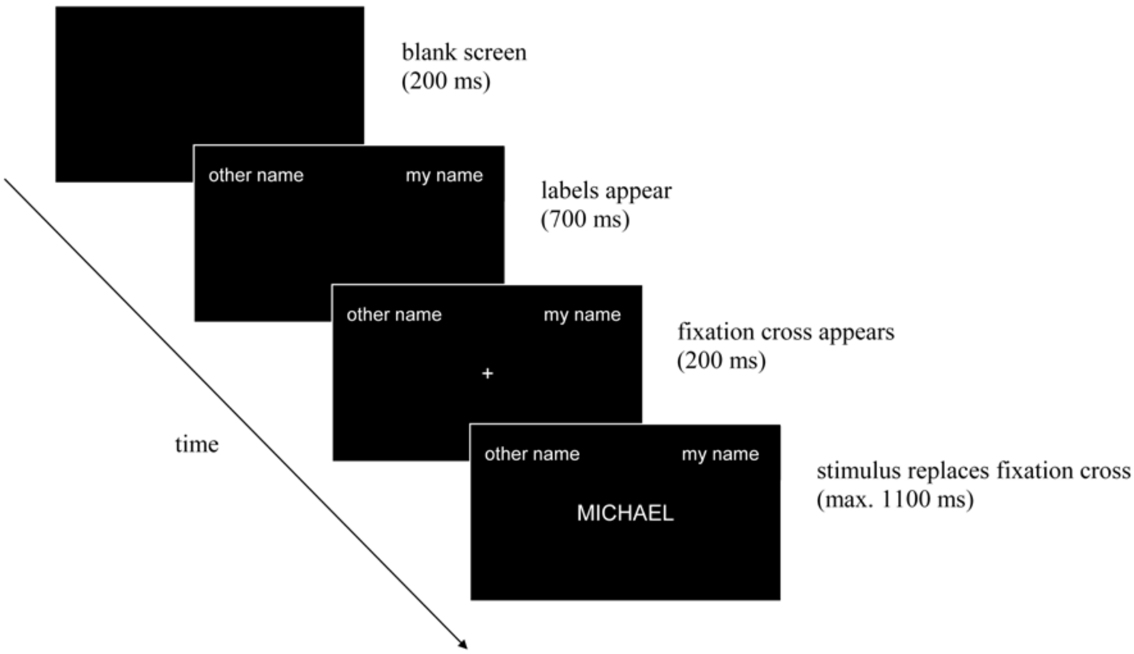

Figure 1. Example of a trial in the A-CIT. First the labels appear, and then follow the stimuli. The stimulus is either an expression referring to own name or an actual given name (including the participant’s own name in the guilty condition). The next trial begins again with a blank screen, and the subsequent labels either appear at the same locations as on the previous trial or they switch positions.

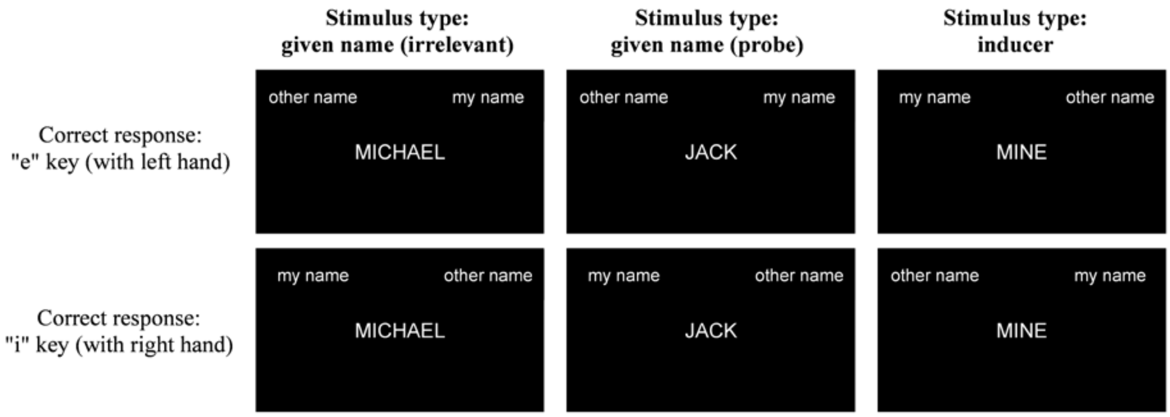

Figure 2. Examples of the possible stimulus type and label position variations in the A-CIT for a participant called 'Jack'. Note that the stimuli are presented completely intermixed during the task, and the labels switch or do not switch places at random. Thus, on each trial, any of these variations may come up – consequently, the participant has to constantly pay close attention to both the labels and the following stimuli. Please note that the presentation and the required response for the probe is exactly the same as for any of the irrelevants.

Procedure

In the guilty condition, the critical items consisted of the participant's given name (as probe item), and four other, irrelevant names. In the innocent condition, the critical items consisted of five irrelevant names – however, unbeknownst to the participant, one of these five names was in fact the name of another participant (i.e., the probe item for another participant), which was subsequently used in the statistical analyses as the "presumed probe". Moreover, this set of five names in one participant's innocent condition was the same as the set of five names in the other participant's guilty condition. This was done in order to obtain a well-controlled comparison on the group level. All participants gave their whole name prior to the experiment on an online application sheet, and all probe and irrelevant items for all participants were generated[2] in advance.[3]

The entire task was automatized (PsychoPy in Python; Peirce, 2007)[4], but an experimenter was always present to answer possible questions. Participants were informed about the details of the "lie detection simulation" experiment on an introduction page, where the purpose and the basic rationale of the lie detection test was explained. They were also informed about the two conditions ("guilty," in which case they have to lie about their name, i.e., deny recognizing it; and "innocent," in which case their name does in fact not appear in the test), emphasizing that in either case the simulation requires that they deny recognizing any of the names in the task as their own, and that they want to seem innocent. After having read the information, participants pressed the spacebar to consent and begin the simulation of the lie detection scenario.

In the main task, each trial began with a blank screen for 200 ms. After this, both labels appeared on the upper part of the screen. After another 700 ms (during which the participant processed the arrangement of the labels), a fixation cross appeared in the middle of the screen, for 200 ms, in order to draw the participant's attention to the coming stimulus. Finally, the stimulus appeared in the place of the fixation cross. The participant had 1100 ms to respond to the stimulus. In pilot studies with this response window, error rates averaged around 10%. This strictly short response window, which made the task difficult to perform, was chosen because (1) it forces the examinee to pay close attention and make fast responses

[2] The details of this generation are described in an online appendix (Appendix B) at https://osf.io/k47cg/.

[3] Due to the excluded participant and participants who signed up but did not come to perform the experiment, 7 participants in the innocent conditions task and 7 in the guilty condition task used item sets that were not used for another participant. Nevertheless, in these cases, for probe items in the analyses of the innocent condition, we still used the given names of the participants who were excluded or did not perform the experiment.

[4] The script is available on request from the first author. The main texts (introduction, instructions) are uploaded at https://osf.io/k47cg/ in Appendix A, containing both original (German) and translated (English) versions.

(which a liar may want to avoid if possible, despite the instructions), and (2) it makes it very difficult to manipulate the timing of the responses (i.e., faking: Verschuere et al., 2009).

The display did not change in case of an incorrect response: either the correct answer or the end of the response window was awaited. Feedback was given only when the correct response was not made within the response window ("Too slow!" caption for 400 ms); see Figure 1.

The main task was preceded by two practice tasks. In the first practice task, the response window was longer than in the main task (2100 ms instead of 1100 ms), and feedback was immediately given in case of an incorrect response (“False!” written in red, below the stimulus), while the second task had the same response window as in the main task (1100 ms) and no feedback in case of an incorrect response. In both practice tasks, expressions referring to other people’s names (e.g., “other” or “theirs”) were presented instead of actual given names: four different expressions referring to other people’s names were presented 8 times, and four different expressions referring to the participant’s name were presented 9 times, in random order (thus altogether 17 trials; the original expressions and their English translations can be found in Appendix A at https://osf.io/k47cg/). Otherwise, the two practice tasks were identical to the main tasks. In either practice task, in case of too few correct responses (below 55%) or too many omitted (too slow) responses (over 20%), participants received a corresponding feedback, were reminded of the instructions, and had to repeat the practice task.

This was followed by a final check to ensure that the participant had understood the task. Expressions referring to other people’s names were now replaced by actual given names, and all possible stimuli were presented once in a random sequence: four expressions referring to the participant’s name, and five actual names – these names were either four irrelevant names and the participant’s own name (guilty condition), or five irrelevant names (but including a "presumed probe"; innocent condition). On each trial, same as in the subsequent main task, the "other name" and "my name" labels changed or did not change places at random, and participants had to classify the presented items according to the labels (expressions referring to the participant's name to "my name" and all actual given names to "other name"). In this short task, participants had plenty of time (10 seconds) 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 the task. All participants had to (and did) complete this task correctly two times. This check guaranteed that the eventual differences (if any) between the responses to the probe item and the responses to the irrelevant items were not due to misunderstanding of the instructions or any uncertainty about the required responses in the eventual task.

The following main task consisted of three blocks of 137 trials, including 80 with actual names (each of the five names 16 times), and 57 with expressions referring to own name (14 times the same four expressions as in the practice task, plus one randomly chosen as the first trial of the block); thus altogether 411 trials in the main task. All stimuli were presented in random order, but with several restrictions (to avoid word repetition and to balance the changing of label positions and stimulus categories).[5] There were breaks between the blocks – participants could take a rest and continue when they felt ready.

For the second A-CIT (for the other condition) the procedure was exactly the same, except that the first practice task was omitted. Participants completed the whole experiment

[5] The same stimulus was never repeated on consecutive trials. The label placement (i.e., "my name" on the left and "other name" on the right, or "my name" on the right and "other name" on the left) was never repeated on more than three consecutive trials. Each given name (the probe, and the four irrelevants) was preceded, in 50% of its appearances, by another given name, and in the other 50% of its appearances, by an expression referring to the participant's own name. Furthermore (and also within each of the two cases described in the previous sentence), each given name was accompanied by the two possible label positions on equal number of trials (i.e., 50% one label position, 50% the other). The expressions referring to the participant's own name were, on average, also accompanied by the two possible label positions on equal number of trials (excluding the first, randomly chosen trial of each block).

(including instructions, the two A-CITs, and debriefing) in 35-40 minutes from their arrival (within this, one full A-CIT took 12-14 minutes).

Data Analysis

Overall rates of correct responses were used to detect outliers in case of responses to personal names, and in case of responses to self-referring expressions. For all subsequent analyses, responses below 150 ms RT were excluded. For RT analyses, only correct responses were used. Accuracy was calculated as number of correct responses divided by number of all trials (after the exclusion of those with an RT below 150 ms).

Along with the conventional values reported for paired-sample t-tests, we also report within-subject Cohen's d values following the formula given in recent RT-CIT studies (Kleinberg & Verschuere, 2015, 2016; Verschuere & Kleinberg, 2015; Verschuere, Kleinberg, & Theocharidou, 2015; adopted from Lakens, 2013), for the sake of comparison between studies.

To assess the efficiency of discriminating between guilty and innocent conditions, we calculated areas under the receiver operating characteristic curve (AUROC curve, or simply AUC – area under the curve; a diagnostic efficiency measure, for binary classification, that takes into account the distribution of all predictor values (see e.g., Zou, O'Malley, & Mauri, 2007). The AUC can range from 0 to 1, where .5 means chance level classification, and 1 means flawless classification (i.e., all guilty and innocent classifications can be correctly made based on the given predictor variable, at a given cutoff point). RT-CIT studies usually use mean RTs and accuracies as the basis of predictor variables. More precisely, they use the difference between the mean RT to probes and the mean RT to irrelevant items, and the difference between the accuracy rate to probes and accuracy rate to irrelevant items, calculated for each individual (e.g., Seymour et al., 2000; Verschuere, Crombez, Degrootte, & Rosseel, 2010; Visu-Petra, Miclea, & Visu-Petra, 2012). Given the complexity of this novel

A-CIT task and the longer response window (compared to the regular RT-CIT), we expected high variability and a skewed distribution of RTs, and therefore we also added a third predictor, median RT – which is, compared to mean RT, less sensitive to outliers and skewness (e.g., Ratcliff, 1993, pp. 522, 531).

We used an alpha level of .05 for all statistical significance tests.

Results

As noted in the Participants section, one participant was found to have an outlier error rate (only 70.8% correct responses in case of personal names) and was excluded from further analyses. The mean rate of correct responses for the remaining participants was 89.9±5.4% for names, and 87.1±6.2% for self-referring expressions.

The results data for the experiment can be retrieved from the Open Science Framework data repository via https://osf.io/ k47cg / (Open Science Collaboration, 2012).

Group-level analysis

All means and SDs of individual RT means, medians, and response accuracies, for the different stimuli types, in guilty and innocent conditions, are given in Table 1.

Table 1. Means and standard deviations of individual reaction time means, medians, and response accuracies, for the different types of stimuli in Experiment 1 and Experiment 2

	Experiment 1		Experiment 2	
	Innocent	Guilty	Innocent	Guilty
Means (ms)				
All names	600±73	607±80	643±110	674±97
Probe	593±74	639±93	643±112	710±97
Irrelevant	601±73	600±78	643±109	665±99
Self-referring	615±78	630±78	651±108	687±97

Medians (ms)				
All names	577±77	590±83	626±118	663±111
Probe	568±76	626±94	627±118	704±100
Irrelevant	580±79	582±81	626±119	651±112
Self-referring	598±82	617±84	639±116	675±103
Accuracies (%)				
All names	90.4±4.9	89.4±5.9	91.2±3.8	90.0±5.8
Probe	90.9±5.8	84.9±9.0	90.9±4.9	85.7±8.1
Irrelevant	90.3±5.2	90.5±5.5	91.2±3.8	91.1±5.9
Self-referring	87.9±6.1	86.2±6.3	89.6±4.6	88.9±5.6

Note. Means and standard deviations (in the format of MEAN±SD) for individual mean RTs, median RTs, and accuracies (percentages of correct responses) for *All names* (including both probe and irrelevant), *Probe* (item presumed to be the participant's own given name), *Irrelevant* (other names), *Self-referring* (expressions referring to own name). The two conditions: Guilty – in which case the Probe was actually the participant's own name; and Innocent – in which case the Probe was not the participant's own name. Unlike in Experiment 1, participants in Experiment 2 were informed about the selected probe item prior to the task (in both guilty and innocent conditions).

To examine the differences between the mean RTs to the probe and those to the irrelevant, and their possible interactions across the two conditions, we performed a repeated-measures ANOVA with the within-subject factors of Type (probe or irrelevant) and Condition (guilty and innocent). The main effect of Type indicated slower responses for probes ($F(1, 26) = 13.6$, $p = .001$, $\eta_p^2 = 0.343$), while the Condition had no significant main effect ($p = .126$).

Most importantly to the present hypotheses, the significant Type x Condition interaction (F(1, 26) = 28.1, p < .001, η_p^2 = 0.519) indicated that the probe-irrelevant difference was larger in the guilty condition. Consequently, to examine whether the main effect of Type was only due to a robust difference in the guilty condition, simple effects were examined. Follow-up paired-sample t-tests indeed revealed that the difference was only significant in the guilty condition (t(26) = 5.17, p < .001, d = 0.995), and not in the innocent condition (t(26) = -1.97, p = .059, d = -0.380). Furthermore, to follow-up the significant Type x Conditon interaction, we also tested the simple effects of Condition, which was found significant regarding probes, i.e., slower responses to probes in the guilty condition, compared to the innocent condition (t(26) = 3.16, p = .004, d = 0.608), while there were no significant differences regarding RTs to irrelevant stimuli (p > .9). Finally, we also compared the two conditions by computing the simple individual differences between probe and irrelevant mean RTs for each condition; i.e., probe mean RT minus irrelevant mean RT calculated for each individual. These probe-irrelevant differences were significantly larger in the guilty than in the innocent condition (t(26) = 5.30, p < .001, d = 1.020).

To examine the differences between the rates of correct responses to probes and those to the irrelevant items, and their possible interactions across the two conditions, the same repeated-measures ANOVA was performed. The main effect of Type indicated lower accuracy to probes (F(1, 26) = 20.4, p < .001, ηp2 = 0.439), and the main effect of Condition indicated lower accuracy in the guilty condition (F(1, 26) = 8.1, p = .008, ηp2 = 0.238). The Type x Condition interaction showed that the probe-irrelevant accuracy difference was larger in the guilty condition (F(1, 26) = 12.0, p = .002, ηp2 = 0.315). Follow-up t-tests revealed that the significant Type main effect was due to significantly lower probe accuracy, compared to irrelevant accuracy, only in the guilty condition (t(26) = 5.05, p < .001, d = 0.972), but not in the innocent condition (p > .5). Furthermore, the effect of Condition was only significant regarding probes, i.e., low accuracy to probes in the guilty condition, compared to the innocent condition ($t(26) = 3.48$, $p = .002$, $d = 0.670$), while there were no such differences regarding accuracies to irrelevant stimuli ($p > .8$). When comparing the two conditions in respect of the simple individual differences between probe and irrelevant accuracies (i.e., irrelevant accuracy minus probe accuracy for each individual), these differences were again significantly larger in the guilty condition ($t(26) = 3.46$, $p = .002$, $d = 0.666$).

The probe-irrelevant differences in mean RT, median RT, and accuracy were not influenced by the main effect of the Order of conditions ($p > .1$ for each measure) or by the Condition x Order of conditions interaction ($p > .1$ for each measure).

For self-referring expressions, mean RTs and accuracies did not differ significantly between the two conditions ($p > .1$ for all paired-sample t-test comparisons).

Individual classification

Probe-irrelevant differences in mean RTs, median RTs, and accuracies were used as predictor variables to calculate AUCs (see Methods, Data Analysis). The AUC was .838 (CI: .722 - .954) for mean RTs, .867 (CI: .761 - .973) for median RTs, and .794 (CI: .674 - .913) for accuracies (see left panel in Figure 3).

In addition, we computed a logistic regression with guilty/innocent as the outcome predicted from the two variables. Assessment of goodness-of-fit revealed a significant improvement relative to a constant-only model (X2(2, N = 54) = 31.444, $p < .001$; Nagelkerke's R2 = .589). The probability of guilty was significantly associated with response time (B = 38.71, Wald X2(1) = 9.785, $p = .002$) and accuracy (B = 18.02, Wald X2(1) = 7.968, $p = .024$). This reflects that both predictors individually contribute to the probability of the outcome guilty. The AUC for the model-based predicted probability of "guilty" was .888 (CI: .802 - .973).

We assessed the generalizability of the model-based classification to new cases using leave-one out cross-validation (LOOCV, Efron & Tibshirani, 1994). In an iterative procedure, we estimated the logistic regression model for N – 1 cases (calibration set), and computed the predicted outcome probability for the remaining case (generalization set). ROC-curve and AUC was then determined for the predicted outcome probabilities across all cases. The corresponding ROC-curve with AUC = .857 (CI: .756 - .959) is shown in Figure 3. The optimal threshold for classification according to the Youden-Index (point on the ROC-curve furthest from the diagonal) was at a predicted probability for the outcome guilty of .39. With this cutoff, 23 out of the 27 participants in the guilty condition were correctly classified as guilty (true positive rate: .85), and 5 out of the 27 were incorrectly classified as guilty in the innocent condition (false positive rate: .19).

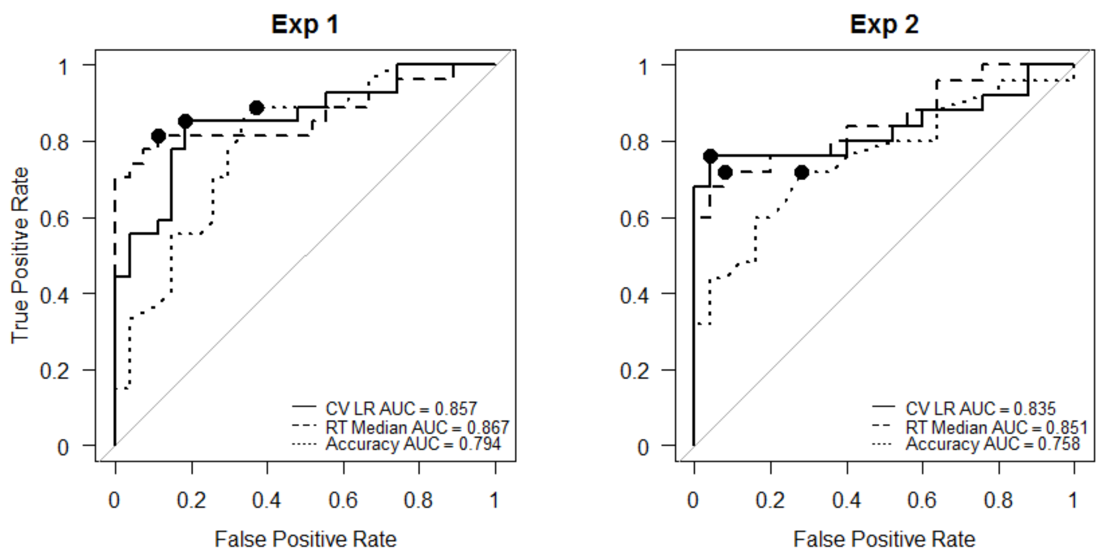

Figure 3. ROC curves for (1) reaction time (RT) medians, (2) accuracies, and (3) probabilities for the outcome guilty from cross-validated logistic regression (CV LR) in Experiments 1 and 2. True positive rates (guilty participants correctly classified as guilty) as a

function of false positive rates (innocent participants incorrectly classified as guilty) using all possible cutoff points. Bold points reflect optimal cut-offs according to the Youden-Index.

Discussion

In this first experiment, participants in the guilty condition responded to probe items significantly more slowly, and with less accuracy, in comparison to the irrelevant items – while no such differences were found in the innocent conditions. This difference between the two conditioned lead to efficient guilty/innocent classifications, showing that the A-CIT is capable of providing high deception detection accuracy. Consequently, a second experiment was run to see whether our paradigm is also resistant to information leakage. The study design was the same as in Experiment 1, except that all participants were informed about the probe item (as a simulation of information leakage) in both guilty and innocent conditions.

Experiment 2

Methods

Participants

Another 28 bachelor students at the Department of Psychology, University of Klagenfurt, Austria volunteered and participated in the experiment. Data from three of these participants were excluded from all analysis due to not recalling the specified probe item at the end of the experiment. This left 25 participants (age = 24.28±5.91 years; 5 male). Fifteen participants were randomly assigned to first perform the A-CIT in guilty condition, and then the A-CIT in innocent condition, while 10 were assigned to perform the two tasks in the reverse order.

Procedure

As in Experiment 1, all irrelevant items for all participants were generated in advance, with the names in each participant's guilty condition used in another participant's innocent condition.[6]

The following procedure replicated Experiment 1, except for the important modification that participants were informed about the probe (or presumed probe) item prior to each of the two A-CITs (i.e., in both conditions). Following the introduction page, participants were presented a brief background story about a person, named e.g., Robin, who committed a serious (unspecified) crime, and who is hiding under false identity. The participant was informed that he/she is one of our suspects, and he/she will be tested to see whether his/her name is actually Robin. Depending on the first condition, the name in the background story was either the participant's own name (probe item; guilty condition) or an irrelevant name (presumed probe item; innocent condition). This name was written four times in different sentences on this page, so that the participant would certainly remark it. The rest of the first A-CIT followed as in Experiment 1. Before the second A-CIT, another background story was presented, which was simply a paraphrased version of the first background story and with another name (probe or presumed probe item, depending on the second condition).

At the end of the experiment (i.e., after both A-CITs were done and the participant was informed that the lie detection simulation is over), the participant was prompted, in a pop-up window, to type in the name that appeared in the one of the two background stories in which it was not his/her own. As noted in the Participants section, three participants were excluded for not remembering the correct name.

Results

[6] Due to the excluded participants and participants who signed up but did not come to perform the experiment, 5 participants in the innocent conditions task and 5 in the guilty condition task used item sets that were not used for another participant.

The mean of overall rate of correct responses was 90.6±4.0% for names, and 89.3±4.2% for self-referring expressions, with no outliers in either case. For all subsequent analyses, responses below 150 ms RT were excluded. The analysis procedure was the same as in Experiment 1.

Same as for Experiment 1, the results data for Experiment 2 can be retrieved via https://osf.io/ k47cg /.

Group-level analysis

All means and SDs of individual RT means, medians, and response accuracies, for the different stimuli types, in guilty and innocent conditions, are given in Table 1.

To examine the differences between the mean RTs to the probes and those to the irrelevant items, and their possible interactions across the two conditions, we again performed a repeated-measures ANOVA with the within-subject factors of Type (probe or irrelevant) and Condition (guilty and innocent). The main effect of Type indicated slower responses for probes ($F(1, 24) = 29.8$, $p < .001$, ηp2 = 0.554), while the main effect of Condition indicated slower responses in the guilty condition ($F(1, 24) = 7.4$, $p = .012$, ηp2 = 0.235). The Type x Condition interaction showed that the probe-irrelevant difference was larger in the guilty condition ($F(1, 24) = 22.3$, $p < .001$, ηp2 = 0.481). Follow-up t-tests revealed that the significant Type main effect was due to significantly slower responses to probes, compared to RTs to irrelevant stimuli, only in the guilty condition ($t(24) = 5.68$, $p < .001$, $d = 1.136$), but not in the innocent condition ($p > .9$). Furthermore, the effect of Condition was only significant regarding probes, i.e., slower responses to the probe in the guilty condition, compared to the innocent condition ($t(24) = -3.86$, $p = .001$, $d = -0.772$), while there were no such differences regarding the mean RTs obtained for irrelevant items ($t(24) = -1.32$, $p = .198$, $d = -0.265$). The individual differences between probe and irrelevant mean RTs were significantly larger in the guilty condition ($t(24) = 4.72$, $p < .001$, $d = 0.944$).

Another repeated-measures ANOVA was performed to compare accuracies for probe and irrelevant items across the two conditions. Again, the main effect of Type indicated lower accuracy to probes (F(1, 24) = 9.7, p = .005, ηp2 = 0.289), and the main effect of Condition indicated lower accuracy in the guilty condition (F(1, 24) = 5.1, p = .033, ηp2 = 0.175). The Type x Condition interaction showed that the probe-irrelevant accuracy difference was larger in the guilty condition (F(1, 24) = 15.9, p = .001, ηp2 = 0.398). Follow-up t-tests revealed that the significant Type main effect was due to significantly lower accuracies to probes, compared to irrelevant items only in the guilty condition (t(24) = -3.97, p = .001, d = -0.794), but not in the innocent condition (p > .7). Furthermore, the effect of Condition was only significant regarding probes, i.e., low accuracies to probes in the guilty condition, compared to the innocent condition (t(24) = 3.454, p = .002, d = 0.691), while there were no such differences regarding accuracies to irrelevant names (p > .9). The individual differences between probe and irrelevant accuracies were significantly larger in the guilty condition (t(24) = 3.45, p = .002, d = 0.691).

The probe-irrelevant differences in mean RT, median RT, and accuracy were not influenced by the main effect of the Order of conditions (p > .2 for each measure) or by the Condition x Order of conditions interaction (p > .1 for each measure).

In the case of self-referring expressions: mean RTs and accuracies did not differ significantly between the two conditions (p > .1 for all paired-sample t-test comparisons).

Individual classification

Same as in Experiment 1, we used probe-irrelevant differences in mean RTs, median RTs, and accuracies as predictor variables. The AUC was .811 (CI: .683 - .939) for mean RTs, .851 (CI: .743 - .959) for median RTs, and .758 (CI: .622 - .893) for accuracies (see right panel in Figure 3). Each of these AUCs in Experiment 2 was compared to the AUC using the same given predictor (mean RTs, median RTs, or accuracies) in Experiment 1, but no significant differences were found (p > .6 for all comparisons using z tests; (Hanley & McNeil, 1982).

As in Experiment 1, we predicted the outcomes guilty/innocent based on response time and accuracy differences using logistic regression. The goodness-of-fit test against a constant-only model was statistically reliable ($X2(2, N = 50) = 27.507$, $p < .001$; Nagelkerke's $R2 = .564$). The probability of the outcome guilty was significantly associated with response times ($B = 44.886$, Wald $X2(1) = 9.586$, $p = .002$), but not with accuracy ($B = 13.663$, Wald $X2(1) = 3.037$, $p = .081$). The model-based AUC was .867 (CI: .761 - .974).

As before, LOOCV was used to test the generalizability of the model-based classification. For comparability with Experiment 1 we included both predictors in the logistic regression model. The AUC of the cross-validated predictions was .835 (.710 - .960). According to the Youden-Index the optimal cut-off was at a predicted probability of .61 for the outcome guilty. At the cut-off, the true positive rate was .76 and the false positive rate .04.

We assessed the generalizability of the cut-offs by classifying cases in Experiment 2 based on the cut-off from the cross-validated logistic regression in Experiment 1. In the guilty condition, 19 out of the 25 participants were correctly classified as guilty (true positive rate: .76), whereas in the innocent condition 8 out of the 25 participants were incorrectly classified as guilty (false positive rate: .32). The results support the validity of the A-CIT, however given that optimal cut-offs and classification performance will vary across samples, other approaches to establish generalizable and robust classification thresholds should be tested in future research.

General discussion

In the present paper, we have introduced a new deception detection method, the A-CIT: an RT-based task that makes use of the natural associations between examinee-related critical items and phrases describing ownership. We have shown, in two independent experiments, that using this method, guilty and innocent conditions can be efficiently differentiated based on differences between the responses to the probe item (i.e., the participant's own name) and the responses to the irrelevant items (i.e., other names): in the guilty condition, the responses to the probe items were slower, and more often incorrect, than the responses to the irrelevant items. Furthermore, in the second experiment, participants were always informed about the probe item prior to the testing (as a simulation for leaked crime details), and yet, the A-CIT's classification efficiency remained high. It is noted that both RT and accuracy measures gave slightly worse results in this second experiment (AUCs between .75 and .86 in Experiment 2, while between .79 and .87 in Experiment 1), but these differences are negligible.

Based on the most efficient predictor (RT medians), we could discriminate between guilty and innocent participants with an AUC of .87 and .85 (in Experiments 1 and 2, respectively), which are fairly high rates considering that a recent meta-analysis found the weighted average of AUCs in RT-CIT studies to be .82 (Meijer, Verschuere, Gamer, Merckelbach, & Ben-Shakhar, 2016). Moreover, and quite importantly, we used a single-probe protocol, i.e., only one type of items (given names). Verschuere et al. (2015) have shown that substantially better accuracies can be obtained using a multiple-probe protocol, i.e., several item types randomly intermixed within the same task (e.g., names, birthdates, nationalities, etc.; see also Eom, Sohn, Park, Eum, & Sohn, 2016). For one, it is quite possible that the A-CIT could also be improved with the inclusion of several item types. For another, there are scenarios in which a single-probe protocol would be preferable or even the only viable option (e.g., when only a single relevant crime detail is known).

Notable differences from the autobiographical Implicit Association Test

Compared to the A-CIT, the main difference is that the aIAT does not use multiple items, but, as noted in the Introduction, only two opposing possibilities (e.g., having or not having used cocaine; Sartori et al., 2008, p. 774). Furthermore, while all items are randomly intermixed in the A-CIT, in the aIAT the critical autobiographical items fixedly alternate with the inducers (i.e., every second trial is an inducer).

Firstly, this makes the aIAT method straightforward and intuitive in structure, giving itself easily to manipulation (e.g., Fiedler & Bluemke, 2005; Röhner, Schröder-Abé, & Schütz, 2013), which was also shown to reduce accuracy below chance level when used for deception detection (Verschuere et al., 2009). Moreover, this faking can be learned by anybody by training oneself using one of the abundant freely available online IAT tasks that also give feedback about the participant's performance. We have not yet tested the resistance of the A-CIT to countermeasures, but, given its complexity, it is very likely to be less susceptible to faking than the aIAT. It is also less likely to be widely available to the public, and therefore practicing countermeasures would be less feasible.

Secondly, in the aIAT, the examinee will always be aware of the relevant question (e.g., whether he/she used cocaine). Studies have shown that this could lead to a false-positive classification, if an innocent examinee just imagines that he/she is guilty (Shidlovski, Schul, & Mayo, 2014; Takarangi, Strange, & Houghton, 2015). The A-CIT may have similar shortcomings when the probe is known to the examinee (this also await further research), but this method can also be used in scenarios where the probe is not known to the examinee – in which case it would function similarly to the original CIT, and would avoid the possibility of such false-positives. In addition, it would also be possible to use the A-CIT in scenarios where the probe is unknown even to the investigators (e.g., the location of an upcoming terrorist attack), and multiple options are presented to find out which of the items is associated with the most deviant (e.g., slowest) responses – which will then be assumed to be the probe (Rosenfeld, 2011, p. 83). A further option in this case is to sequentially narrow the array of possibilities to find the answer – e.g., first locating the country, then the city, etc. This would require a single-probe protocol, at which the original RT-based CIT does not perform well.

Finally, the aIAT would be somewhat more difficult to standardize for widespread use in different situations (and different languages) because it uses full sentences as items – while in the A-CIT, only simple words (or very short expressions) have to be provided.

Future research

The A-CIT method, as presented in the present paper, leaves many possibilities for improvements that could increase its accuracy rates even further. For one, continually switching the positions of the labels might result in substantial statistical noise in the data, which would decrease the classification accuracy of the task. This “switching” could be replaced by other methods that increase attention to the meaning of the labels (e.g., the Extrinsic Affective Simon Task, De Houwer & De Bruycker, 2007; or the Brief Implicit Association Test, Sriram & Greenwald, 2009; see also: Krause, Back, Egloff, & Schmukle, 2011). However, we also note that the constant attention to unexpectedly switching labels imposes a high cognitive load to the participants, which has been repeatedly shown to be beneficial in detecting concealed information (e.g., Visu-Petra, Varga, Miclea, & Visu-Petra, 2013).

The basic parameters of the task (e.g., the ratios of the different categories, the inter-stimulus intervals, the randomization process, etc.) were optimized during numerous pilot tests, but – same as in the case of other RT-CITs – they could be tested more extensively and thoroughly in the future. For practical purposes, it may be an asset to use an extended practice block procedure to calibrate the duration of the response window individually.

In our study, given names were the objects of the test, but the task can very easily be generalized. Most evidently, the object could be any other autobiographical detail, e.g., place of origin or birthday – in which case the labels would be e.g., “my birthday” and “other birthday," while the self-referring expressions would stay the same, except that of course "my name" would again be replaced by "my birthday." Moreover, the same principle could just as well work in case of a crime, e.g. for a murderer's gun ("my gun") or for a stolen object ("my loot"). We acknowledge that this design may have limitations, since e.g., a thief might not consider a stolen object as his/her own property. However, in future research, the validity of action related expressions as inducers (replacing ownership related expressions) could also be explored, e.g., "I stole", "they stole", etc. A further option is phrases depicting ownership of actions (e.g., "I did") as inducers and action verbs as critical items (e.g., "steal").

Finally, the A-CIT could easily be combined with other deception detection methods that use sequentially presented simple stimuli (e.g., polygraph, EEG). Using the same or a similar task, the focus on the associations may not only lead to larger differences in RT responses, but may also improve the differentiability of the physiological responses to the probe item (e.g., larger electrodermal responses or larger P300 waves).

The validity of the A-CIT in correctly classifying cases as guilty is promising and should be further tested in direct comparison to other deception detection methods as well as in innocent and guilty scenarios that more closely reflect the conditions of real-life investigations.

Author Contributions

Gáspár Lukács conceived, designed, and conducted the experiment, performed most of the statistical analyses, and prepared the manuscript. Bartosz Gula gave advice and helped in connection with the implementation and conduction of the experiment, performed some of the statistical analyses, reviewed and wrote some parts of the manuscript. Emese Szegedi-Hallgató helped with the programming of the experimental software. Gábor Csifcsák oversaw and gave advice on the experiment design, reviewed and wrote some parts of the manuscript.

Acknowledgements

We are very grateful to Dávid Farkas and several others who repeatedly pretested the experimental task. Special thanks to Dorota Goc, who, besides pretesting, proofread the first draft of the manuscript.

References

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

Ben-Shakhar, G. (2012). Current Research and Potential Applications of the Concealed Information Test: An Overview. *Frontiers in Psychology*, *3*. http://doi.org/10.3389/fpsyg.2012.00342

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*(3), 214–234. http://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. http://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: Cambridge University Press.

De Houwer, J., & De Bruycker, E. (2007). The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(2), 133–143. http://doi.org/10.1016/j.jbtep.2006.10.004

Efron, B. & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. London: Chapman & Hall.

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. http://doi.org/10.14257/ijmue.2016.11.2.23

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and Unaided Response Control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*(4), 307–316. http://doi.org/10.1207/s15324834basp2704_3

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.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. http://doi.org/10.1148/radiology.143.1.7063747

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. http://doi.org/10.1037/a0023589

Kleinberg, B., & Verschuere, B. (2015). Memory Detection 2.0: The First Web-Based Memory Detection Test. *PLOS ONE*, *10*(4), e0118715. http://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. http://doi.org/10.1016/j.jarmac.2015.11.004

Krause, S., Back, M. D., Egloff, B., & Schmukle, S. C. (2011). Reliability of implicit self-esteem measures revisited. *European Journal of Personality*, *25*(3), 239–251. http://doi.org/10.1002/per.792

Kraut, R. (1980). Humans as Lie Detectors. *Journal of Communication*, *30*(4), 209–218. http://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*. http://doi.org/10.3389/fpsyg.2013.00863

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: Procedures and validity. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59-102). New York: Guilford Press.

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

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty: Deception research: Methodological considerations. *Psychophysiology*, n/a-n/a. http://doi.org/10.1111/psyp.12609

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45–69. http://doi.org/10.1037/a0030734

Open Science Collaboration. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(6), 657–660. https://doi.org/10.1177/1745691612462588

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

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).

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. http://doi.org/10.1037/0033-2909.114.3.510

Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions. *Journal of Research in Personality*, *47*(4), 330–338. http://doi.org/10.1016/j.jrp.2013.02.009

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

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the Implicit Association Test: The Recoding-Free Implicit Association Test (IAT-RF). *The Quarterly Journal of Experimental Psychology*, *62*(1), 84–98. http://doi.org/10.1080/17470210701822975

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to Accurately Detect Autobiographical Events. *Psychological Science*, *19*(8), 772–780. http://doi.org/10.1111/j.1467-9280.2008.02156.x

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. http://doi.org/10.1037//0021-9010.85.1.30

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*(2), 314–335. http://doi.org/10.1037/0033-295X.115.2.314

Shidlovski, D., Schul, Y., & Mayo, R. (2014). If I imagine it, then it happened: The Implicit Truth Value of imaginary representations. *Cognition*, *133*(3), 517–529. http://doi.org/10.1016/j.cognition.2014.08.005

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*(4), 283–294. http://doi.org/10.1027/1618-3169.56.4.283

Takarangi, M. K. T., Strange, D., & Houghton, E. (2015). Event familiarity influences memory detection using the aIAT. *Memory, 23*(3), 453–461. http://doi.org/10.1080/09658211.2014.902467

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. http://doi.org/10.1016/j.sbspro.2014.03.335

Verschuere, B., Crombez, G., Degrootte, T., & Rosseel, Y. (2010). Detecting concealed information with reaction times: Validity and comparison with the polygraph. *Applied Cognitive Psychology*, *24*(7), 991–1002. http://doi.org/10.1002/acp.1601

Verschuere, B., & Kleinberg, B. (2015). ID-Check: Online Concealed Information Test Reveals True Identity. *Journal of Forensic Sciences*, n/a-n/a. http://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. http://doi.org/10.1016/j.jarmac.2015.01.001

Verschuere, B., & Meijer, E. H. (2014). What's on Your Mind?: Recent Advances in Memory Detection Using the Concealed Information Test. *European Psychologist*, *19*(3), 162–171. http://doi.org/10.1027/1016-9040/a000194

Verschuere, B., Prati, V., & Houwer, J. D. (2009). Cheating the Lie Detector: Faking in the Autobiographical Implicit Association Test. *Psychological Science, 20*(4), 410–413. http://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, *Deception detection: Current challenges and new approaches* (pp. 269–291). Oxford, UK: John Wiley & Sons.

Visu-Petra, G., Miclea, M., & Visu-Petra, L. (2012). Reaction Time-based Detection of Concealed Information in Relation to Individual Differences in Executive Functioning: RT-based CIT and executive functioning. *Applied Cognitive Psychology, 26*(3), 342–351. http://doi.org/10.1002/acp.1827

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*. https://doi.org/10.3389/fpsyg.2013.00146

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation*, *115*(5), 654–657. http://doi.org/10.1161/CIRCULATIONAHA.105.594929