Nationality check in the face of information contamination

testing the Inducer-CIT and the autobiographical IAT

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

DOI
10.1080/1068316X.2022.2102170

Publication date
2024

Document Version
Final published version

Published in
Psychology, Crime and Law

License
CC BY

PSYCHOLOGY, CRIME & LAW
2024, VOL. 30, NO. 5, 458–477
https://doi.org/10.1080/1068316X.2022.2102170

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Nationality check in the face of information contamination: testing the Inducer-CIT and the autobiographical IAT

Dave Koller [a,b], Franziska Hofer [c,d], Signe Ghelfi [c] and Bruno Verschuere [b]

[a]Department of Psychology, University of Zurich, Zurich, Switzerland; [b]Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands; [c]Zurich State Police, Airport Division, Research and Development, Zurich, Switzerland; [d]Brainability, Developing Human & Organizational Potentials, Zurich, Switzerland

ABSTRACT
Government agencies including border control have an interest to detect if someone provides false information about their nationality. While response time tasks have been proposed to be able to detect someone's true nationality, there is the risk that they will often err, particularly in the face of information contamination (i.e. someone having thorough knowledge of the country). We screened 2,200 participants to create three groups: Dutch participants ($n = 118$), and British participants with ($n = 99$) and without knowledge of the Netherlands ($n = 118$). They were tested with either the autobiographical Implicit Association Test (aIAT) or the Inducer-Concealed Information Test (I-CIT). While both tests could discriminate Dutch participants from British participants without knowledge of the Netherlands ($AUC_{I\text{-}CIT} = .65$; $AUC_{aIAT} = .88$), only the aIAT could also discriminate Dutch participants from British participants with knowledge ($AUC_{I\text{-}CIT} = .52$; $AUC_{aIAT} = .86$). Therefore, the aIAT, but not the I-CIT, could be a helpful tool to detect false nationality claims, even when information contamination is suspected.

ARTICLE HISTORY
Received 20 May 2021
Accepted 24 June 2022

KEYWORDS
Concealed Information Test; autobiographical Implicit Association Test; memory detection; deception detection; reaction time

Introduction

Lying about nationality

The Covid-19 pandemic forced the world to drastic measures. In Europe, the pandemic challenges the core principles of the European Union: solidarity, policy coordination, and especially the free movement across national borders (Biancotti et al., 2020). Being able to travel between countries and continents both due to available means of transportation and international policy agreements is one major achievement of modern human society. However, the motivation to travel is not always due to innocent or voluntary nature. For example, human trafficking, international terrorism, and structural criminality are some of the negative concomitants of globalization. According to the report of the

CONTACT Dave Koller d.koller@psychologie.uzh.ch Department of Psychology, University of Zurich, Binzmühlestrasse 14 / Box 22, Zurich 8050, Switzerland

United Nations, the number of international migrants worldwide reached nearly 272 million in 2019, of which about 24 million are considered refugees and asylum seekers (United Nations, 2019). Between 2010 and 2017, the number of refugees and asylum seekers increased by about 13 million (World Bank, 2019). The process for asylum is clearly regulated in the law. In the case of Switzerland, the Asylum Act states the possibilities to get asylum. Furthermore, in most cases the application for asylum must be submitted directly at the border. Usually, according to the Swiss State Secretary of Migration an interrogation is conducted about the travel itinerary, family background, and other aspects. However, the time for the final decision can take years (Roos et al., 2018; Swiss State Secretary of Migration, 2020). To this end, the person seeking asylum has most likely endured an exhausting, maybe even dangerous journey, and is often not allowed to take a job. If the application for asylum is rejected, the person may need to go back to the home country independent of the time spent abroad (Federal Act on Foreign Nationals and Integration, 2005). Consequently, whether due to criminal intention or despair, one might be motivated to provide false information about the own nationality and other autobiographical aspects (e.g. travel itinerary, already applied asylums, family background) to increase the chance of getting the application for asylum approved. In other words, people might lie about their true nationality to be permitted to stay in the country or to be allowed to take a job. Let us take a look at the following example: Due to the humanitarian catastrophe in Eritrea (European Asylum Support Office, 2019), asylum applications were very likely to be approved (European Asylum Support Office, 2015). This was not the case for the neighboring country Ethiopia. However, Ethiopia is also facing political instability (e.g. Pilling & Schipani, 2020). Therefore, for someone with Ethiopian origin suffering from hunger and having lost hope of a better future, it could be attractive to seek asylum by falsely claiming to be from Eritrea. As the person might be able to tell a lot about Eritrea – perhaps because of thorough research, family history, or because of having lived in Eritrea – it could be feasible to pretend to have an Eritrean origin and hide their true nationality. On the other hand, a person from Eritrea might mistakenly be accused to be from Ethiopia and rejected for asylum. So, the question is: is it possible to detect the true nationalities?

Detecting lies about nationality

In general, it is a well-replicated fact that humans are not very good at detecting whether someone is telling the truth or not (on average 54% accuracy; e.g. Bond & DePaulo, 2006, 2008; Hartwig & Bond, 2011). Specific techniques, as for example those arising from the cognitive approach of lie detection (e.g. imposing cognitive load, encouraging to say more, or asking unanticipated questions; see Vrij, 2015 for a review) can be used in order to amplify verbal and nonverbal differences between liars and truth tellers. In a meta-analysis, Vrij and colleagues assessed the effectiveness of this approach to be around 71% correct classification (Vrij et al., 2017). While that seems to be an improvement compared to the 54% base rate, the authors' analysis has been criticized by Levine et al. (2017) stating that the true detection rate of the cognitive load approach is much lower, and the benefit was overestimated. Moreover, the most recent meta-analysis points out that the benefit of the cognitive approach is reduced remarkably when considering publication bias (Mac Giolla & Luke, 2020).

Apart from novel interviewing techniques, researchers also worked on technology-assisted ways to detect deception. A well-investigated test is the reaction time (RT)-based Concealed Information Test (CIT) originally named as Guilty Knowledge Test (Lykken, 1959; Seymour et al., 2000; for a review see Verschuere & De Houwer, 2011). After Kleinberg and Verschuere (2015) showed the RT-CIT can be run reliably and validly via the internet, this test has also gained attention by practitioners. The test allows to identify concealed knowledge, i.e. it can detect if somebody has specific knowledge about a topic of interest (e.g. a country, crime, name, etc.). In the RT-CIT, a series of stimuli is presented on the computer screen. There are three types of stimuli: the item of interest (so-called probes), irrelevant items that have no specific relevance to the participants nor are connected to the case, and targets. Targets are used to make sure that participants are paying attention and processing the stimuli. For the Eritrea-Ethiopia asylum case, an RT-CIT could be operationalized in the following way: the probes could be names of towns, locations, mountains, or products of Ethiopia that are mainly known by indigenous people and rather not known by non-indigenous. Target items could be names of towns, locations, mountains, or products of Eritrea. Irrelevants could be names of rather unknown towns, locations, mountains, etc. of other eastern African countries. The question used in the RT-CIT would then be 'Is this connected to your home country?'. Participants would then be instructed to respond with 'Yes' to the targets (names of towns, locations, mountains, or products of Eritrea) and with 'No' to all other items. People with an Ethiopian origin recognize the probes and therefore show longer RTs for the probes than for the irrelevant due to response conflict (Seymour & Schumacher, 2009; Verschuere & De Houwer, 2011). Thus, recognition is inferred based on systematically longer RTs to Ethiopian items (i.e. the probes) as compared to the irrelevant stimuli (i.e. towns in other African countries).

Information contamination: a challenge to detect lying about nationality

Research shows that the RT-CIT constitutes a reliable and valid method of detecting concealed recognition and the potential applications are manifold. The meta-analysis of 114 studies on 3307 participants by Suchotzki et al. (2017) showed a large effect ($d = 1.049$; 95% CI [0.930; 1.169]). One major concern, the RT-CITs vulnerability to faking, was recently addressed by Suchotzki et al. (2021). In a series of studies, they showed that faking was ineffective when participants only had a short response window. However, another major restriction is that the RT-CIT and the physiological CIT are not immune against information contamination, meaning that validity is threatened if the critical information (probe) could also be recognized by innocent examinees (Lukács & Ansorge, 2019; see Bradley et al., 2011 for a review). For example, a person from Eritrea seeking asylum could know a lot about Ethiopia, e.g. because he/she has worked or lived in the country, has heard news about Ethiopia in the press, or just because of personal interest. In this case, the person would show longer RTs to the probe although not being Ethiopian.

Depending on the context, this so-called information contamination could happen quite often. Consequently, lack of immunity of the CIT against information contamination is one of the most often mentioned concerns of practitioners (see Podlesny, 2003; and interviews conducted by the authors with Swiss law enforcement offices of different branches). Theoretically, there are at least two ways to tackle this challenge. In the context of a crime, information contamination could be mitigated by a very concise information management by the investigative authority. However, for cases such as the mentioned asylum example, this possibility seems less feasible as due to globalization, digitalization, and social media information are distributed worldwide. For example, a person stating to be from a specific country has many possibilities to gain as much knowledge as possible about this country. He or she could learn online about the politics, could walk on the streets in the capital city with Google maps, and could follow people on social media twittering about daily events. Keeping information within a restricted group of people poses an increasing challenge in the globally networked world. Therefore, the second possibility to reduce the negative effect of information contamination could be achieved by further developing the RT-CIT method to increase the method's robustness against information contamination.

Possible solutions to information contamination

In this regard, Lukács and Ansorge (2019) recently proposed an adjusted RT-CIT. The I-CIT utilizes a shared feature (e.g. home-relatedness) between the so-called inducer items and the probes to induce response conflict. To illustrate the principle, we come back to our Eritrea-Ethiopia asylum example. In the I-CIT the participant is asked to categorize home-referring inducers (e.g. 'HOME', 'NATIVE'), foreign-referring inducers (e.g. 'FOREIGN', 'OTHERS'), and various names of Ethiopian mountains, products, and towns. The name of the Ethiopian items and the foreign-referring inducers share the same response key. This is expected to lead to response conflict in residents of Ethiopia. For them, Ethiopian items refer to home, but they must press the response key associated with foreign. No response conflict is expected in residents of Eritrea, even if they know the Ethiopian mountains, towns, and products, because they are still foreign. Using autobiographical information, Lukács and Ansorge (2019) showed that a 'guilty' group that has been instructed to hide their identity showed the expected I-CIT-effect. Moreover, an 'informed innocent' group (i.e. non-deceptive participants who possess the critical information) did not show the I-CIT-effect, suggesting the I-CIT is immune to information contamination.

The I-CIT bares some similarity to the autobiographical Implicit Association Test (aIAT; Sartori et al., 2008), an established test for assessing autobiographical memory. With the aIAT it is possible to evaluate which of the two autobiographical events is true (for a review see Agosta & Sartori, 2013). True events are identified by shorter RTs when sharing the same response key as the 'true' category compared to when the event shares the response key with the 'false' category. Verschuere and Kleinberg (2017) showed that the aIAT can also be used successfully in web-experiments for assessing autobiographical memory. While the aIAT has applied potential, it must be taken into account that a meta-analysis found it had lower average effect size compared to the CIT (Suchotzki et al., 2017). One explanation could lie in an important design difference: The RT-CIT varies trials with and without response conflict on a trial by trial basis, whereas the aIAT varies them by blocks. Participants may therefore more easily control their RTs in the aIAT compared to the CIT.

The current study

The goal of this study was to examine the impact of information contamination on the I-CIT and the aIAT. To our knowledge, this is the first study that investigates information contamination on the aIAT. This was done with an online experiment using a scenario of identifying the true nationality. For the study, Dutch and British participants were recruited. Within the British participants, there were two groups: The information-contamination group consisted of British people which all had good knowledge about factual details of the Netherlands. The other group consisted of British people without any specific knowledge about the Netherlands. Therefore, these three groups resulted: Dutch group, UK naïve group, and UK contamination group. Our main predictions were that the I-CIT and the aIAT show above chance classification performance both when discriminating the Dutch from the UK naïve as well as when discriminating the Dutch from the UK contamination group. We further test the hypotheses that the predictors of the I-CIT (*dCIT*) and the aIAT (D_1) are larger in the UK contamination group than in the UK naïve group which would be a sign against immunity to information contamination. Finally, we investigated if there is a difference in the Dutch/UK classification performance between the I-CIT and the aIAT.

Method

The experiment was approved by the ethical committee of the Faculty Social and Behavioral Sciences of the University of Amsterdam (approval number: 2020-CP-2352). Data, code, and preregistration are publicly available on https://osf.io/8wyf6/.

Deviations from preregistration

After data collection, but before testing the hypotheses, we realized that one of our exclusion criteria (the recognition test) was not well designed and would have led to the unnecessary exclusion of the majority of our sample (64–76%, for the aIAT and I-CIT respectively). Stage 1 (screening) served to assess eligibility for our study, and those who met inclusion criteria were invited for stage 2 (the RT-tasks). At the end of stage 2, we had a brief recognition test that we planned to use as a double check of group assignment. In this test, participants were presented with 3 Dutch (probe) items amongst 12 (irrelevant) foils and participants should mark the Dutch items. However, we did not set a limit to the number of items they could select nor have any penalty for wrong responses. Participants selected on average 4.7 items, hence most made at least one 'error'. Sticking to the preregistered exclusion criterion would entail excluding most participants based on a dubious test. We therefore dropped this measure as ground for exclusion. We do report the results of the planned, but underpowered, analyses and show our main findings hold.

Procedure

Stage 1 – screening. Dutch and British participants that were between 18 and 50 years old were eligible to participate in the screening. This took participants about three minutes to complete, and participants were compensated with 0.35 GBP (≈ 0.50 USD). Participants were recruited over the online platform Prolific.

Participants provided informed consent. Participants were asked to indicate the country that they would call their 'home' and their 'native' country, to estimate the amount of time they spent in the Netherlands in their entire life and to rate their knowledge about the Netherlands on a 7-point Likert scale (1 = very poor, 7 = very good). Then, participants were presented with ten single-choice knowledge questions about the Netherlands in which participants needed to indicate which of the five options (e.g. Zeldonk, Ipelo, Utrecht, Winddicht, Omert) is connected to the Netherlands. For the complete list of items, see https://osf.io/8wyf6/. To discourage participants to look up the answers, we did not use incentives for correct answers and each question needed to be answered within 10 s. After every question, participants were asked to indicate their confidence that the answer is correct on a 7-point Likert scale (1 = not confident at all; I guessed, 4 = relatively confident, 7 = extremely confident).

Because we were interested in the knowledge of participants, we required a question to be answered correctly and with a confidence of 5 or higher to be scored as 'known'. Based on the answers and confidence ratings, we looked for the three questions that divide the screened participants most evenly into the following three groups while including as many of the screened participants as possible. Group 1 (Dutch): Participants who indicated the Netherlands as their home and native country, and who knew the answers to those questions about the Netherlands whose items will be used in the I-CIT. Group 2 (UK contamination): Participants who indicated the United Kingdoms as their home and native country, and who that also knew the answers to those questions about the Netherlands. Group 3 (UK naïve): Participants who indicated the United Kingdoms as their home and native country, and who answered the questions about The Netherlands incorrectly or with a confidence of 1 (i.e. guessing).

Stage 2 – RT-task. In a second stage, one to three days after the screening, participants were invited to the RT-task. The possibility to participate in the second stage ended eight days after the invitation. Participants again provided informed consent.

Participants were asked to imagine a scenario in which they need to use an online service that is not available for Dutch citizens. During the registration for this service, they indicated to be from the United Kingdoms. Participants were told that the service provider is required to use precautions against people providing wrong nationality information and that the following task is used to verify if they truly are from the United Kingdoms. Participants should try to convince this automated test to be from the United Kingdoms. This was the truth for the UK naïve and UK contamination group, and a lie for the Dutch group. Then, the task specific instructions followed, which differed depending on the assignment to either the I-CIT or the aIAT task. Task assignment was done based on the participant number which was randomly assigned by Inquisit (version 6.2.1; Inquisit, 2020) – the software used for the RT-tasks.

Inducer-Concealed Information Test. The I-CIT was mimicked after Lukács and Ansorge (2019). Participants were asked to answer the question 'Is this connected to my home?'. They were presented with the three YES-inducer items ('Home', 'Native', 'Local') and six NO-inducer items (e.g. 'Foreign', 'Abroad', 'Others'). Participants were told that, to be convincing, they need to press YES ('A'-key) only when one of the YES-inducer items appeared. In all other cases they need to press NO ('L'-key). Figure 1 shows an exemplary segment of the I-CIT. The complete list of items used in the I-CIT is displayed in Table 1.[1] They were also instructed to respond as fast and accurately as possible. Feedback was given throughout the task by a 'TOO SLOW' appearing on top of the stimulus after

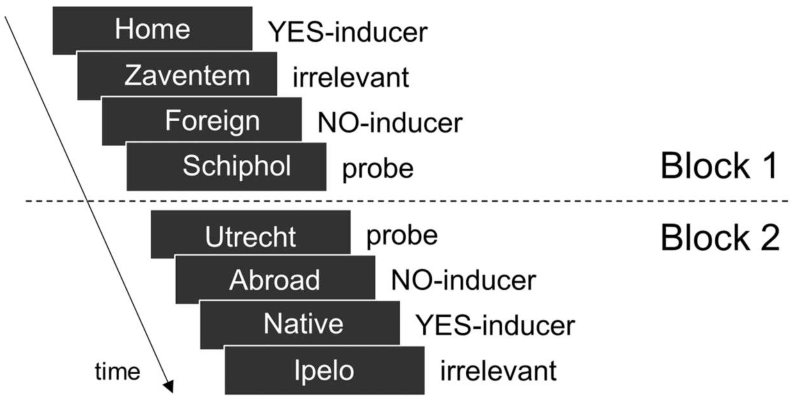

Figure 1. Exemplary segments of the I-CIT.

Table 1. Items used in the I-CIT.

Item category	Items		
	Location	City	Club
Probe	Schiphol	Utrecht	Ajax
Irrelevant	Zaventem, Kortrijk, Ondermeer, Oostende	Winddicht, Omert, Ipelo, Zeldonk	Cres, Teucer, Eetion, Priam
YES-Inducer	Home, Native, Local		
NO-Inducer	Foreign, Abroad, Others, Theirs, Another, Alien		

800 ms until a response was given or the response deadline of 1500 ms was reached, and by a 'WRONG' message appearing below the stimulus for 400 ms in case of an error. Participants completed a practice block with 24 trials in which each item was presented once in random order. If participants had less than 50% correct for any item type, the practice block was repeated (up to two times). The main task contained three blocks, one for each information (Location, City, Club) with a short, self-paced break between blocks. Each block started with 8 burn-trials that were excluded from the analysis to get participants accustomed to the task again after the break. Then followed the 135 test trials that were divided into three sub-blocks of 45 trials (5 YES-inducer, 10 NO-inducer, 6 probes, 24 irrelevants). Item order within a sub-block was randomized.[2] Every inducer was presented 5 times, and every irrelevant or probe item 18 times per block. This results in a total of 405 test trials (54 probe, 216 irrelevant, 45 YES-inducer, 90 NO-inducer).

Autobiographical Implicit Association Test. The scenario in the aIAT condition was the same. Participants were asked to use the 'A' and 'L'-keys to categorize statements into response categories shown on the top left and top right of the screen. Statements that were presented in green (logical statements) needed to be categorized into TRUE or FALSE; statement presented in white (autobiographical statements) into 'I am from the Netherlands' and 'I am from the United Kingdoms' (see Table 2 for the complete list of statements).[3] Participants were instructed to answer as fast and accurately as possible. Feedback was given throughout the task by a 'WRONG' message appearing below the statement until the correct response was given (following the D_1 scoring scheme of

Table 2. Statements used in the aIAT.

Category		Statements
Logical	True	I am in front of a screen
		I am participating in a study
		I am reading words
		I am classifying sentences
		I am using a keyboard
	False	I am climbing a mountain
		I am in the sea
		I am looking at birds
		I am watching a movie
		I am using a pan
Autobiographical	The Netherlands	The Netherlands is my home country
		I am native to the Netherlands
		My nationality is Dutch
		I own a Dutch Passport
		I come from the Netherlands
	United Kingdoms	The United Kingdom is my home country
		I am native to the United Kingdom
		My nationality is British
		I own a British Passport
		I come from the United Kingdom

Greenwald et al., 2003). The aIAT is divided into seven blocks with brief instructions at the beginning of each block. The logical discrimination block (block 1, 20 trials) consisted of logical statements only. The initial autobiographical discrimination block (block 2, 20 trials) consisted of autobiographical statements only. In the initial double categorization practice block (block 3, 20 trials), and in the initial double categorization block (block 4, 60 trials) participants alternately saw logical and autobiographical statements and needed to classify them both according to their respective, color coded, categories. The reversed autobiographical discrimination block (block 5, 20 trials) consisted of autobiographical statements only, but with reversed category-key bindings compared to block 2. Finally, the reversed double categorization practice block (block 6, 20 trials) and the reversed double categorization block (block 7, 60 trials) combined the category-key bindings from block 1 and block 5. The category-key bindings of block 2 were assigned based on the subject number which was randomly generated by Inquisit. Data from blocks 3, 4, 6, and 7 are used in the analysis (see D_1 measure in Greenwald et al., 2003).

Post-test recognition test. After the RT-task, participants were asked to select the items from the list of all probes and irrelevants for which they are ‘fairly certain to very certain’ that they are connected to the Netherlands.[4] Finally, participants were thanked, debriefed, and given the code to be entered on Prolific.

Participants

After screening the preregistered maximum of 2200 participants (300 Dutch, 1900 British), we had 209 participants in the Dutch group, 155 in the UK contamination group, and 260 in the UK naïve group, for a total of 624 participants eligible for the RT task. Of those, 355 (57%) completed the second stage. Twenty participants (6%) were excluded from the analysis. One participant revoked the consent, 19 participants (8 Dutch, 2 contamination, 9 naïve) did not reach the pre-registered minimal requirement of at least 40% correct on each item type in the I-CIT.[5] No participant needed to be excluded in the aIAT condition.

Table 3. Demographic information.

		Dutch	Contamination	Naïve	Collapsed
I-CIT	*n*	45	44	64	153
	Sex	33% female	20% female	81% female	49% female
	Age	26.8 (7.4)	34.5 (7.7)	27.7 (8.4)	29.4 (8.5)
aIAT	*n*	73	55	54	182
	Sex	34% female	36% female	70% female	46% female
	Age	25.9 (7.2)	34.2 (8.0)	28.1 (7.5)	29.1 (8.3)
Collapsed	*n*	118	99	118	335
	Sex	34% female	29% female	76% female	47% female
	Age	26.3 (7.3)	34.3 (7.8)	27.9 (8.0)	29.2 (8.4)

Note: Standard deviations are reported in the parentheses.

The final sample consisted of $n = 335$ participants (54% of the invited participants). It took participants about 10 min to complete the RT-task which was reimbursed with 1.10 GBP (≈ 1.50 USD). Demographic information per condition is shown in Table 3.

Results

Following the preregistration, we excluded incorrectly answered trials and trials with latencies smaller than 200 ms or larger than 1499 ms from the analysis in the I-CIT. A total of 3480 trials (5.6%; 2695 inducer trials, 785 non-inducer trials) were excluded. In the aIAT, we only excluded trials with latencies larger than 10,000 ms as preregistered and suggested by Greenwald et al. (2003). A total of 6 trials (0.02%) were excluded.

To classify participants in the I-CIT condition, we used the normalized within participant probe-irrelevant difference $dCIT = \frac{M_{RT}(probe) - M_{RT}(irrelevant)}{SD_{RT}(irrelevant)}$ (Kleinberg & Verschuere, 2015). Dutch participants are expected to show a positive *dCIT* score indicating recognition of the probes and a link between the probe and the participants' home, according to Lukács and Ansorge (2019). Naïve participants are expected to show a *dCIT* score around zero indicating that the probe could not be distinguished from irrelevant items. If the I-CIT indeed is immune to information contamination (Lukács & Ansorge, 2019), then we expect *dCIT* scores of the contamination group to be around zero (because the probe is not linked to the participants' home); if not, we expect positive *dCIT* scores (due to probe recognition).

Individual classification in the aIAT was done using the D_1-measure proposed by Greenwald et al. (2003). D_1 is an equal weight average between $D_{Practice}$ and D_{Test}. ($D_1 = \frac{D_{Practice} + D_{Test}}{2}$). $D_{Practice}$ and D_{Test} are the differences in mean RT between the congruent versions (the labels 'TRUE' and 'I am from the United Kingdoms' share the response key) and incongruent versions (the labels 'TRUE' and 'I am from the Netherlands' share the response key) of the respective block divided by their pooled standard deviation (e.g. $D_{Test} = \frac{M_{RT}(congruent\ testblock) - M_{RT}(incongruent\ testblock)}{SD_{Pooled}}$, with $SD_{Pooled} = \sqrt{\frac{(n_{congruent} - 1)*var(RT_{congruent}) + (n_{incongruent} - 1)*var(RT_{incongruent})}{n_{congruent} + n_{incongruent} - 2}}$). D_1 values greater than zero result from slower responding in congruent blocks than in the

Table 4. Mean reaction times.

Group	I-CIT Probe	I-CIT Irrelevant	I-CIT Probe-Irrelevant Difference	aIAT True/ 'I am from the UK' and False/ 'I am from the Netherlands'	aIAT False/ 'I am from the UK' and True/ 'I am from the Netherlands'	aIAT Block Difference
Dutch	442 (50)	438 (50)	4.2 (15.8)	1026 (183)	962 (216)	64 (166)
UK contamination	426 (45)	423 (43)	3.2 (11.9)	823 (159)	983 (224)	−159 (185)
UK naïve	406 (46)	409 (45)	−3.4 (11.7)	887 (213)	1087 (213)	−199 (211)

Note: RTs are reported in ms, standard deviations are reported in parentheses.

incongruent blocks (i.e. a response conflict in the congruent block). Therefore, positive D_1 values are expected for Dutch participants. Negative D_1 values on the other hand indicate British nationality.

While we rely, per our preregistration, solely on the *dCIT* and D_1 scores for hypothesis testing, we report RTs for both tasks in Table 4 to provide a comprehensive picture of how participants of the three conditions responded in the two tasks. Inspection of Table 4 shows larger probe-irrelevant differences for knowledgeable participants (Dutch, UK contamination) than for naïve participants in the I-CIT. Table 4 further shows that the Dutch group had longer RTs for the TRUE/ 'I am from the United Kingdoms' block than for the TRUE/ 'I am from the Netherlands' block. Both UK groups showed the reversed pattern.

To help assess the evidential value of the seemingly small between group differences in the I-CIT, we conducted a Bayesian analysis of variance with a Cauchy prior (scale = .5) of group (Dutch, UK contamination, UK naïve) on the probe-irrelevant difference. This analysis was not preregistered and is not used for hypothesis testing. The data were 7 times more likely under the hypothesis that there is an effect of group on the probe-irrelevant difference than under the null-hypothesis without a group effect ($BF_{10} = 7.0$). We also conducted post-hoc group comparisons using a two-sided Bayesian t-test with a Cauchy prior (scale = .707). The uncorrected Bayes factors indicating how many times more likely the data is under the hypothesis that there is a difference between the respective groups relative to the null-hypothesis were: $BF_{Dutch/Naïve,0} = 7.2$, $BF_{Dutch/Contamination,0} = .23$, and $BF_{Naïve/Contamination,0} = 7.3$. The same Bayesian analysis for the aIAT showed the data was $4*10^{11}$ times more likely under the hypothesis that there is an effect of group relative to the null-hypothesis ($BF_{10} = 4.0*10^{11}$). The results of the individual post-hoc comparisons show that the data were more likely under the alternative hypothesis (i.e. that there is a between group difference) relative to the null-hypothesis for Dutch vs. UK naïve and Dutch vs UK contamination comparison but not for the UK naïve vs UK contamination comparison ($BF_{Dutch/Naïve,0} = 3.9*10^9$; $BF_{Dutch/Contamination,0} = 1.4*10^8$; $BF_{Naïve/Contamination,0} = .33$).

Preregistered analyses

Without information contamination (i.e. discrimination between Dutch and UK naïve participants), the classification performance of the I-CIT was slightly above chance with an area under the receiver operator characteristics curve (AUC) of .65 (95% CI: [.55, .76]). In other words, the I-CIT can distinguish between Dutch and naïve British participants above chance, but only slightly. The critical classification in a scenario with information contamination, however, is between the Dutch and the UK contamination group. The I-

CIT could not distinguish between the two groups (AUC = .52; 95% CI: [.40, .64]) which is not significantly different from chance level performance.

Using D_1 for the same classifications in the aIAT conditions showed above chance performance, not only for the Dutch versus UK naïve classification (AUC = .88; 95% CI: [.82, .94]), but also for the critical Dutch versus UK contamination classification (AUC = .86; 95% CI: [.80, .93]). These results show that not only can the aIAT also successfully distinguish between Dutch and knowledgeable British participants, it can do so with a performance comparable to the Dutch–naïve classification.

While successful Dutch versus UK contamination classification is crucial for applied purposes, this is not sufficient to show that a test protocol is immune to information contamination. If a test protocol is truly immune to information contamination, then the index used to classify a person should not be influenced by information contamination. A Bayesian one-sided independent samples *t*-tests with Cauchy priors (scale = .707) showed that the data are 17.8 times more likely under the hypothesis that the *dCIT* score for the contamination group is larger than that of the naïve group, compared to the null (or immunity) hypothesis that they do not differ. Thus, the I-CIT is clearly not immune to information contamination (see Figure 2). A Bayesian one-sided independent samples t-tests with Cauchy priors (scale = .707) showed that the data are 3.45 times more likely under the null (or immunity) hypothesis that the D_1 scores of the aIAT do not differ between the contamination group and the naïve group, compared to the null hypothesis that the scores of the contamination group would be larger than those of the naïve group. Thus, there was some evidence that the aIAT is immune to information contamination (see Figure 3).

Finally, we directly compared the classification performance of the I-CIT and the aIAT in the task that resembles the problem practitioners face the most (i.e. to identify the Dutch participants among all three groups). We calculated DeLongs' test for differences in ROC curves. The classification performances differed significantly, $D(226.6) = 4.65$, $p < .001$, with the aIAT showing more accurate classification (AUC = .87; 95% CI: [.82, .92]) than the I-CIT (AUC = .60; 95% CI: [.50, .70]).

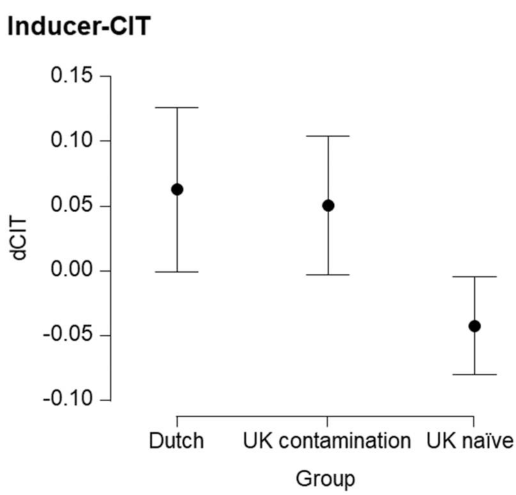

Figure 2. Mean I-CIT scores (dCIT). Note: The 95% credible interval indicates the range in which the true parameter falls with a probability of 95%.

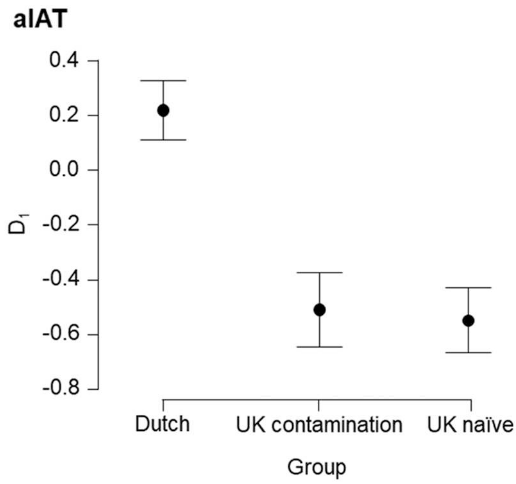

Figure 3. Mean aIAT scores (D_1). Note: The 95% credible interval indicates the range in which the true parameter falls with a probability of 95%.

Exploratory analyses

To make sure that our deviation from the preregistration did not drastically change the results and conclusions, we also tested the main hypotheses on the reduced data set (i.e. after exclusions based on the post-test recognition; leaving $n = 103$ divided into $n_{\text{I-CITdutch}} = 10$, $n_{\text{I-CITcontamination}} = 6$, $n_{\text{I-CITnaïve}} = 21$, $n_{\text{aIATdutch}} = 20$, $n_{\text{aIATcontamination}} = 23$, and $n_{\text{aIATnaïve}} = 23$). The only result that qualitatively changed was that the I-CITs performance of discriminating Dutch and naïve UK participants was not above chance in the reduced data set (AUC = .58; 95% CI: [.33, .82]). Note that this result was based on only 31 participants (10 Dutch, 21 naïve UK). The aIAT showed high classification performance also in the reduced data set (AUC = .87; 95% CI: [.77, .98]). Regarding the Dutch versus UK contamination classification, the I-CIT did not perform above chance (AUC = .53; 95% CI: [.21, .85]) but the aIAT did (AUC = .82; 95% CI: [.70, .95]), confirming the results of the main analysis.

Discussion

Key aims and findings

Computerized paradigms such as the RT-CIT have been shown to be able to detect that someone provides false information about their nationality. However, assessing recognition of details related to nationality, there is a risk the test will err for truth tellers who are knowledgeable about the tested nationality. The present study addressed this information contamination scenario, which is problematic for the traditional RT-CIT (Lukács & Ansorge, 2019). However, two other RT-based paradigms might not be affected by information contamination: the novel I-CIT (Lukács & Ansorge, 2019) and the aIAT (Sartori et al., 2008). We tested Dutch participants, and British participants with and without knowledge about the Netherlands. We were primarily interested to see how well the two tests would perform in the most challenging situation: to discriminate Dutch participants from British participants with knowledge about the Netherlands (i.e. contamination group). Our results showed that the aIAT was able to discriminate between the Dutch and the contamination group, but the I-CIT was not.

The I-CIT is vulnerable to information contamination

Contrary to the results of Lukács and Ansorge (2019), we found a larger I-CIT-effect for the UK contamination group than for the UK naïve group, suggesting the I-CIT is vulnerable to contamination. In their original study introducing the I-CIT, Lukács and Ansorge (2019) used autobiographical information (country of origin, date of birth, and favorite animal) and familiarity-related inducers (e.g. 'Familiar', 'Recognized', 'Mine', 'Unfamiliar', 'Unknown', 'Foreign' etc.). The inducers indicating familiarity required a YES response, all other stimuli including the probes a NO response. For guilty participants, they argued that because familiar inducers and the probes share the self-relatedness feature, but require a different response, there is a response conflict. This leads to longer response times for probes as compared to the irrelevant items (which are not self-related, hence no conflict). And indeed, they found the guilty group, but not the contaminated innocent group, to show an I-CIT effect. In contrast, we found I-CIT effects (of similar magnitude) both in the Dutch and the UK contamination group which is a strong indication that home-relatedness did not cause the effect. We see at least two possible reasons for this discrepancy: our implementation to test for home-relatedness was suboptimal or home-relatedness is not a suitable feature for the I-CIT. First, we tested for the nationality whereas Lukács and Ansorge (2019) tested for the identity of participants. As a result, we used home-relatedness as the conflict inducing feature and adapted the inducer items accordingly (incorporating the feedback from the original paper's first author). It could be, however, that the probes (Schiphol, Ajax, and Utrecht) were not strongly connected to 'home' for all participants. For instance, for someone typically relying on the local airport instead of Schiphol. If this is true for a substantial part of Dutch participants, this could explain why did not find larger I-CIT effects for the Dutch than the UK contamination group. Second, home-relatedness might simply not be a suitable feature to induce response conflict in the I-CIT. Due to the lack of research on the I-CIT, we can only speculate about what properties a feature needs to reliably lead to an I-CIT effect (e.g. high saliency).

The validity of the I-CIT

The I-CIT did not differentiate between the Dutch and UK contaminated group, but we did find a small I-CIT effect in both groups. If the response conflict was not induced by home-relatedness, what mechanism might explain this small I-CIT effect? We see at least two possibilities. First, our choice of irrelevant items could be argued to have caused an I-CIT-effect. We used fictional and lesser-known names as irrelevant items. This led to a feature (word familiarity) shared between the probes and the inducers. While this feature is not only shared with the YES-but also with the NO-inducers (since these are also words), some degree of response conflict or at least response uncertainty could have influenced the Dutch and the UK contamination group. Second, the probes were recognized as task relevant by the Dutch and the UK contamination group. This made the probes stand out among a majority of irrelevant items which could lead to an orienting response (Lykken, 1974; Sokolov, 1963) disrupting the decision-making and increasing the RTs (Verschuere et al., 2004; for a review see Verschuere & Ben-Shakhar, 2011). The slowing due to the orienting response is relatively small but comparable to the RT differences reported in our study. The two explanations are not mutually exclusive. But in either case response conflict due to home-relatedness cannot explain the observed I-CIT effects.

Verschuere and De Houwer (2011) argued that stimulus-response incompatibility and the resulting response conflict is crucial to find robust probe-irrelevant differences. In the classic RT-CIT, target items are used to manipulate the stimulus-response compatibility of probes depending on the participant's knowledge. Naïve participants can perform the task solely by judging the familiarity of items since only the targets are familiar to them. For knowledgeable participants, probes are also familiar, but they are instructed to press the key related to unfamiliarity leading to stimulus-response incompatibility. Suchotzki et al. (2018) showed that increased target familiarity and more targets lead to larger probe-irrelevant differences. They argued that participants relied more on familiarity to do the RT-CIT. Matsuda et al. (2009) omitted target items entirely and did not find a probe-irrelevant differences in RTs (while differences in the event related potentials persisted). That is not to say that the familiarity of targets is the only way to induce response conflict. Lukács and Ansorge (2019) argued that other features (e.g. self-relatedness) also lead to response conflict if the feature is shared by the YES-inducers and the probe. However, of the three YES-inducers that they used only 'MINE' was self-referring – 'FAMILIAR' and 'RECOGNIZED' both referred to familiarity. It is therefore unclear why these should not have induced response conflict in the contamination group as well. This, in addition to our findings, warrants caution and further investigation about the assumptions and boundary conditions of the I-CIT before applied use should be considered.

The aIAT as a tool to detect false nationality claims

The aIAT accurately discriminated between Dutch and British participants and did not show reduced classification performance for British participants with knowledge about the Netherlands. This is plausible because the aIAT is not based on the recognition of information but on the associations between the response labels (Sartori et al., 2008). In blocks in which the autobiographic label 'I am from the Netherlands' is paired with the logical label 'True', response times for participants for which the statement 'I am from the Netherlands' is true (the labels are associated; i.e. Dutch participants) will respond faster than in blocks in which 'I am from the Netherlands' is paired with 'False'. Those associations are independent of the participant's knowledge. Our results are in line with previous aIAT studies without a contamination group (for reviews see Agosta & Sartori, 2013; Suchotzki et al., 2017) and show that the aIAT can be a valid tool to assess someone's nationality even if that person has knowledge about the country of his fake nationality.

The good classification accuracy of the aIAT, even in the face of information contamination, does not imply the aIAT is flawless. If participants are instructed on how to fake the aIAT, e.g. by slowing down responses in one block, they are able to do so (Agosta, Ghirardi, et al., 2011; Hu et al., 2012; Suchotzki et al., 2017; Verschuere et al., 2009). Although faking could not be prevented so far, Agosta, Ghirardi, et al. (2011) developed algorithms to detect faking. However, it seems likely that this algorithm could be tricked if the participants get instructed to also slow their responses in the single categorization blocks (blocks 1, 2, and 4). Another limitation of the aIAT regards the labels and statements that can be used. Agosta, Mega, et al. (2011) showed that negative statements (e.g. 'I do not own a Dutch passport'), counter-affirmative statements (e.g. 'I own a passport from another country than the Netherlands'), and negative labels (e.g. 'I am not from the Netherlands') reduced the aIATs classification performance. Therefore, the aIAT needs to contrast two specific autobiographical facts. In a forensic setting, this could be the crime and the alibi. In the context of nationalities, the examiner needs to have a strong suspicion about the true country of origin for the aIAT to perform optimally. Lastly, even though information contamination does not seem to be a problem for the aIAT, association contamination might be. In three mock crime studies, Dhammapeera et al. (2020) showed that the aIAT could accurately distinguish between guilty participants that executed a mock crime and innocent participants that fulfilled a control task. However, the classification performance of the aIAT was reduced when participants imagined the contrasting statement as their alibi A similar problem could arise in the nationality confirmation setting if a British citizen was born and raised in the Netherlands, and views the Netherlands as her/his home. Formally, this person is British but emotionally she/he is Dutch. Research on how to overcome this challenge might be valuable for practical applications. Any imperfect classification system in the context of deception detection raises ethical issues about when it should be applied and how the results should be incorporated in the decision-making process. It could be tempting to strongly rely on such tests, since they do not pose any risk to the examinee, they are easily applicable, and they give the examiner an objective result. However, after carefully validating the test outside of the laboratory, we urge practitioners to use the information provided by tests like these as an addition to other sources of information and not as a replacement.

Limitations

This study is not without its limitations. We see at least three limitations worth discussing. First, we chose a quasi-experimental design to increase the ecologic validity by using pre-existing knowledge about the Netherlands. In terms of internal validity, however, this is inferior compared to a truly experimental design and introduces the possibility of between group differences (e.g. gender, age, level of education, motivation) that could impact the response times other than the nationality and knowledge. For instance, gender happened to covary with group. But an exploratory analysis showed there was no impact on the I-CIT-effect, see supplementary materials (https://osf.io/8wyf6/).

Second, the applied goal of this study also called for a scenario in which such a test might be applied, such as nationality confirmation. While we did our best to optimize both tests for this scenario, it might not be ideal for the I-CIT. Strictly speaking, we did not test nationality directly, but we tested the home-relatedness of the probes. It seems reasonable to assume that the probes are not home related for the two British groups, but it is less clear for the Dutch group. For instance, not all participants may strongly relate the specific probes used (Schiphol, Ajax, and Utrecht) to their home. If and by how much this diminished the I-CIT-effect for the Dutch group cannot be estimated in this study. However, the comparison between the UK naïve and the UK contamination group is not affected by this. A second point connected to the items used in the I-CIT also needs to be addressed. Unexpectedly, we found a slightly negative I-CIT effect for naïve participants, which means that naïve participants responded quicker to the probes than to the irrelevant items. A core assumption of the CIT is that probes and irrelevant items should be indistinguishable by naïve participants and lead to similar RTs. We can only assume that this is due to item properties. However, such item effects reduce the *dCIT* scores for all groups, hence cannot explain that the contaminated groups showed an I-CIT effect.

Third, it is yet to be investigated to what extent our results generalize to other situations. Again, we started out with a problem practitioners face and applied both test to the best or our ability and in consultation of the tests' inventors. For the aIAT, this meant using unambiguous, simple labels and sentences that needed to be classified. One could argue that we did not use the same information in both tests (e.g. to use 'Schiphol' in both tests, we could have used sentences like 'Schiphol is in my home country') to make the aIAT and the I-CIT more comparable in this regard. However, this would be an unnecessarily indirect way of assessing the citizenship with the aIAT and could lead to worse performance. If our results still hold in other scenarios (e.g. a mock crime with 'wallet' as a probe for the I-CIT and the corresponding sentence 'I stole the wallet' for the aIAT) remains to be investigated. It could also be that the home-relatedness feature is not strong enough to induce response conflict (and therefore not suitable to assess the nationality) but the I-CIT would show better performance with identity information (as used by Lukács & Ansorge, 2019) or specific crime knowledge.

Avenues for future research

In future research, it would be valuable to compare the accuracy of RT-based tests to procedures that are currently applied to determine nationality. Lacking hard facts (e.g. a passport, that may have been lost or thrown away) assessing nationality can be a difficult task. Surveying Dutch Immigration Service officers, van Veldhuizen et al. (2017) found that it is common to probe asylum seekers on their knowledge (e.g. home country, the flight route to Europe, etc.). The answers on such questions are then assessed on indicators like consistency and plausibility (Maegherman et al., 2018). But these indicators are poorly specified, and their validity is faint at best (DePaulo et al., 2003). Therefore, it seems valuable to establish the validity of current procedures in a controlled setting where one has ground truth (e.g. a passport), and to compare it with the RT-based tests.

Conclusion

Recognition-based paradigms such as the RT-CIT are error-prone when truth tellers know the to-be tested information. The present study suggests the I-CIT, a variant of the RT-CIT, is no exception. At the same time, our results show that the aIAT was not affected by knowledge about the falsely claimed nationality. Therefore, the aIAT could be a valuable instrument to identify people claiming to have a false nationality – a problem that boarder control and immigration agencies face daily.

Notes

1. Due to changes in the scenario, we could not use the same inducer items as Lukács and Ansorge (2019). However, we contacted the developer of the I-CIT, Gáspár Lukács, to discuss our adjustment before data collection. He agreed that the inducers should be adjusted to the scenario and suggested a minor change to one of the inducer items which we incorporated.
2. Note that this is more random than in the original study by Lukács and Ansorge (2019) in which an inducer was never followed by another inducer.
3. We consulted the developer of the aIAT, Giuseppe Sartori, about the labels and items (logical and autobiographical) and adapted two logical items according to his suggestions.
4. On average, participants selected 4.7 items even though only 3 items were connected to the Netherlands. While this is a sign that we chose realistic items, it also indicates that participants applied a low threshold of certainty to select an item. However, the post-test recognition test can only serve its purpose as a knowledge check when a high certainty threshold is applied. Therefore, we did not use data of the post-test recognition test.
5. While this seems like a low benchmark, there is a strong response tendency towards the 'no' button in the I-CIT because participants need to press 'yes' only for home-referring inducers which make up only 11% of test trials.

Acknowledgements

We also thank Gáspár Lukács and Giuseppe Sartori for their advice on the construction of the Inducer-CIT and the aIAT, respectively.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

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

Data availability statement

The material, data, and scripts that support the findings of this study are openly available on OSF (https://osf.io/8wyf6/).

ORCID

Dave Koller http://orcid.org/0000-0001-5315-8782
Franziska Hofer http://orcid.org/0000-0001-5734-4347
Signe Ghelfi http://orcid.org/0000-0002-9491-2371
Bruno Verschuere http://orcid.org/0000-0002-6161-4415

References

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (2011). Detecting fakers of the autobiographical IAT. *Applied Cognitive Psychology*, *25*(2), 299–306. https://doi.org/10.1002/acp.1691

Agosta, S., Mega, A., & Sartori, G. (2011). Detrimental effects of using negative sentences in the autobiographical aIAT. *Acta Psychologica*, *136*(3), 296–306. https://doi.org/10.1016/j.actpsy.2010.05.011

Agosta, S., & Sartori, G. (2013). The autobiographical IAT: A review. *Frontiers in Psychology*, *4*, 1–12. https://doi.org/10.3389/fpsyg.2013.00519

Biancotti, C., Borin, A., Cingano, F., Tommasino, P., & Veronese, G. (2020, March 18). *The case for a coordinated COVID-19 response: No country is an island*. VoxEU. https://voxeu.org/article/case-coordinated-covid-19-response-no-country-island

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* (pp. 187–199). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.011

DePaulo, B. M., Malone, B. E., Lindsay, J. J., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Dhammapeera, P., Hu, X., & Bergström, Z. M. (2020). Imagining a false alibi impairs concealed memory detection with the autobiographical Implicit Association Test. *Journal of Experimental Psychology: Applied*, *26*(2), 266–282. https://doi.org/10.1037/xap0000250

European Asylum Support Office. (2015). *EASO-Bericht über Herkunftsländer-Informationen: Länderfokus Eritrea*. https://www.easo.europa.eu/sites/default/files/public/BZ0415327DEN.pdf

European Asylum Support Office. (2019). *Eritrea – national service, exit, and return*. https://coi.easo.europa.eu/administration/easo/PLib/2019_EASO_COI_Eritrea_National_service_exit_and_return.pdf

Federal Act on Foreign Nationals and Integration of 16 December 2005, SR 142.20. (2005). https://sherloc.unodc.org/cld/uploads/res/document/federal-act-on-foreign-nationals_html/Federal_Act_on_Foreign_National_EN.pdf

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

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

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

Inquisit 6. (2020). [Computer software]. https://www.millisecond.com/

Kleinberg, B., & Verschuere, B. (2015). Memory detection 2.0: The first web-based memory detection test memory. *PLoS One*, *10*(4), 1–17. https://doi.org/10.1371/journal.pone.0118715

Levine, T. R., Blair, J. P., & Carpenter, C. J. (2017). A critical look at meta-analytic evidence for the cognitive approach to lie detection: A re-examination of Vrij, Fisher, and Blank (2017). *Legal and Criminological Psychology*, *23*(1), 7–19. https://doi.org/10.1111/lcrp.12115

Lukács, G., & Ansorge, U. (2019). Information leakage in the response time-based Concealed Information Test. *Applied Cognitive Psychology*, *33*(6), 1178–1196. https://doi.org/10.1002/acp.3565

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

Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist*, *29*(10), 725–739. https://doi.org/10.1037/h0037441

Mac Giolla, E., & Luke, T. J. (2020). Does the cognitive approach to lie detection improve the accuracy of human observers? *Applied Cognitive Psychology*, *35*, 1–8. https://doi.org/10.1002/acp.3777

Maegherman, E., van Veldhuizen, T. S., & Horselenberg, R. (2018). Dropping the anchor: The use of plausibility in credibility assessments. *Oxford Monitor of Forced Migration*, *7*(2), 37–55. https://www.researchgate.net/publication/331383215_Dropping_the_anchor_the_use_of_plausibility_in_credibility_assessments

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

Pilling, D., & Schipani, A. (2020, November 18). Ethiopia crisis: 'A political mess that makes fathers fight sons'. *Financial Times*. https://www.ft.com/content/b888c23a-45ed-4937-9154-3117cc23e202

Podlesny, J. A. (2003). A paucity of operable case facts restricts the applicability of the guilty knowledge technique in FPI criminal polygraph examinations. *Forensic Science Communications*, *5*(3), 20–37.

Roos, K., Wenger, I., Sowe, R., & Indermühle, Y. (2018). Addressing barriers to work for asylum seekers: Report from Switzerland. *World Federation of Occupational Therapists Bulletin*, *74*(2), 123–127. https://doi.org/10.1080/14473828.2018.1540100

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science*, *19*(8), 772–780. https://doi.org/10.1111/j.1467-9280.2008.02156.x

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

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

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Pergamon Press.

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

Suchotzki, K., Verschuere, B., & Gamer, M. (2021). How vulnerable is the reaction time Concealed Information Test to faking? *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1016/j.jarmac.2020.10.003

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428–453. https://doi.org/10.1037/bul0000087

Swiss State Secretary of Migration. (2020). *Monitoring Asylsystem: Bericht 2019*. https://www.sem.admin.ch/dam/sem/de/data/publiservice/berichte/monitoring-asyl/monitoring-asylsystem-2019.pdf

United Nations. (2019). *International migration 2019: Report*. https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/InternationalMigration2019_Report.pdf

van Veldhuizen, T. S., Maas, R. P. A. E., Horselenberg, R., & van Koppen, P. J. (2017). Establishing origin: Analysing the questions asked in asylum interviews. *Psychiatry, Psychology and Law*, *25*(2), 283–302. https://doi.org/10.1080/13218719.2017.1376607

Verschuere, B., & Ben-Shakhar, G. (2011). Theory of the Concealed Information Test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection. Theory and application of the Concealed Information Test* (pp. 128–148). Cambridge University Press.

Verschuere, B., Crombez, G., & Koster, E. H. W. (2004). Orienting to guilty knowledge. *Cognition and Emotion*, *18*(2), 265–279. https://doi.org/10.1080/02699930341000095

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: Response-latency based measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection. Theory and application of the Concealed Information Test* (pp. 46–62). Cambridge University Press.

Verschuere, B., & Kleinberg, B. (2017). Assessing autobiographical memory: The web-based autobiographical Implicit Association Test. *Memory*, *25*(4), 520–530. https://doi.org/10.1080/09658211.2016.1189941

Verschuere, B., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking in the autobiographical implicit association test: Research report. *Psychological Science*, *20*(4), 410–413. https://doi.org/10.1111/j.1467-9280.2009.02308.x

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Wiley series in the psychology of crime, policing and law*. *Detecting deception: Current challenges and cognitive approaches* (pp. 205–229). Wiley-Blackwell.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1–21. https://doi.org/10.1111/lcrp.12088

World Bank. (2019). *Refugee population by country or territory of asylum*. https://data.worldbank.org/indicator/SM.POP.REFG