



## UvA-DARE (Digital Academic Repository)

### Memory-Based Deception Detection: Extending the Cognitive Signature of Lying From Instructed to Self-Initiated Cheating

Geven, L.M.; Ben-Shakhar, G.; Kindt, M.; Verschuere, B.

**DOI**

[10.1111/tops.12353](https://doi.org/10.1111/tops.12353)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Topics in Cognitive Science

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Geven, L. M., Ben-Shakhar, G., Kindt, M., & Verschuere, B. (2020). Memory-Based Deception Detection: Extending the Cognitive Signature of Lying From Instructed to Self-Initiated Cheating. *Topics in Cognitive Science*, *12*(2), 608-631. <https://doi.org/10.1111/tops.12353>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



This article is part of the topic “Lying,” Hans van Ditmarsch, Petra Hendriks and Rineke Verbrugge (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview)

# Memory-Based Deception Detection: Extending the Cognitive Signature of Lying From Instructed to Self-Initiated Cheating

Linda M. Geven,<sup>a,b</sup> Gershon Ben-Shakhar,<sup>b</sup> Merel Kindt,<sup>a</sup>  
Bruno Verschuere<sup>a</sup>

<sup>a</sup>*Department of Clinical Psychology, University of Amsterdam*

<sup>b</sup>*Department of Psychology, Hebrew University of Jerusalem*

Received 6 September 2017; received in revised form 14 February 2018; accepted 16 April 2018

---

## Abstract

From a cognitive perspective, lying can be regarded as a complex cognitive process requiring the interplay of several executive functions. Meta-analytic research on 114 studies encompassing 3,307 participants (Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017) suggests that computerized paradigms can reliably assess the cognitive burden of lying, with large reaction time differences between lying and truth telling. These studies, however, lack a key ingredient of real-life deception, namely self-initiated behavior. Research participants have typically been instructed to commit a mock crime and conceal critical information, whereas in real life, people freely choose whether or not to engage in antisocial behavior. In this study, participants ( $n = 433$ ) engaged in a trivia quiz and were provided with a monetary incentive for high accuracy performance. Participants were randomly allocated to either a condition where they were instructed to cheat on the quiz (mimicking the typical laboratory set-up) or to a condition in which they were provided with the opportunity to cheat, yet without explicit instructions to do so.

---

Correspondence should be sent to Linda M. Geven, Department of Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, Postbus 15933, 1001 NK Amsterdam, Netherlands. E-mail: l.m.geven@uva.nl

[Correction added on June 26, 2018, after initial online publication: The Topic Editor name in the topic title was changed from “Petra Hendricks” to “Petra Hendriks”.]

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Assessments of their response times in a subsequent Concealed Information Test (CIT) revealed that both instructed cheaters ( $n = 107$ ) and self-initiated cheaters ( $n = 142$ ) showed the expected RT-slowness for concealed information. The data indicate that the cognitive signature of lying is not restricted to explicitly instructed cheating, but it can also be observed for self-initiated cheating. These findings are highly encouraging from an ecological validity perspective.

*Keywords:* Concealed Information Test (CIT); Memory detection; Deception; External validity; Cheating; Honesty

---

## 1. Introduction

### 1.1. A cognitive view on deception

Cognition-based lie detection focuses on the assumption that lying is cognitively more demanding than truth telling. Liars often require a significant amount of mental processes in order to actively suppress the pre-potent truth and appear innocent (Spence et al., 2001; Verschuere, Spruyt, Meijer, & Otgaar, 2011). People report to experience lying as associated with a larger cognitive effort than truth telling (Caso, Gnisci, Vrij, & Mann, 2005; Vrij, Semin, & Bull, 1996). A study tracking participants' arm movements while moving a video-game remote to either a deceptive or truthful option on the screen, confirmed that lying takes longer than telling the truth (Duran, Dale, & McNamara, 2010; Spence et al., 2001; Walczyk, Roper, Seemann, & Humphrey, 2003). Neurocognitive research found that inhibiting the truth and constructing a lie are reflected by increased activation in several regions of the brain associated with cognitive control (e.g., iFG; see Abe, Suzuki, Mori, Itoh, & Fujii, 2007; Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Nunez, Casey, Egner, Hare, & Hirsch, 2005; Sip, Roepstorff, McGregor, & Frith, 2008). These findings inspired a new focus of lie detection techniques that aim to improve differentiation between liars and truth tellers by adding additional cognitive load, such as asking interviewees to tell their story in reversed order (Evans, Michael, Meissner, & Brandon, 2013; Vrij et al., 2008) or by maintaining eye contact (Vrij, Mann, Leal, & Fisher, 2010).

### 1.2. Response time as an index of cognitive effort involved in lying

Renewed attention has drawn researchers to the easily applicable behavioral measure of response time as an indicator for (increased inhibition associated with) deception. Although initial research, often using suboptimal measurement conditions, did not find strong effects of reaction times (RTs) as a deception measure (see Luria, 1932; Marston, 1920), the increased use of computerized measures led to revived research interest.

Seymour, Seifert, Shafto, and Mosmann (2000), for instance, found that RTs provide a fast and reliable indication of recognition of concealed information. Research participants committed a mock crime in the laboratory and were subsequently asked to do a seemingly unrelated binary classification task on the computer. While measuring RTs,

participants had to indicate whether they recognized the stimuli presented in the Concealed Information Test (CIT; Lykken, 1959) by pressing one of two response keys. Critical details from the committed crime were intermixed with irrelevant, neutral items. Upon measuring response latencies on these critical details (*probes*) in comparison to irrelevant words, 89% of the guilty participants could be correctly classified as such.

Another study using a mock-theft scenario, instructed participants to either steal a CD with exam questions or to read a newspaper article about the incident. In a subsequent autobiographical Implicit Association Task (aIAT; Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008), participants had to classify sentences describing possible autobiographical events (e.g., 'I am in front of a computer' or 'I am currently on the beach') to either the 'true' or 'false' label by pressing one of two response keys. All 15 guilty participants made a faster association between the crime-related questions (e.g., 'I stole the CD with the exam questions') and the label 'true', while 13 of the 15 innocent participants were faster combining denials of theft with 'true', leading to a ROC area of 0.96.

Meta-analytic research on 114 studies encompassing 3,307 participants (Suchotzki et al., 2017) using various computerized RT paradigms, shows the potential of RTs to index deception, with a large effect size (Cohen's  $d = 1.05$ ). Yet the results are nearly exclusively based on non-forensic samples and studies lacking key ingredients of real-life deception.

### 1.3. On the limitations of instructed lying and cheating

Deception is commonly defined as a voluntary act (see Vrij, 2004), in which deliberation and intention are key factors. Moreover, Sip et al. (2008) pointed out that deception comes without forewarning and instructions on when to deceive. Yet, in laboratory studies on detecting deception, participants are often explicitly instructed to lie (e.g., Furedy, Davis, & Gurevich, 1988) or commit a staged crime and subsequently conceal knowledge (e.g., Lykken, 1959; Nahari & Ben-Shakhar, 2011). As a consequence, questions can be raised about the suitability of these studies, involving explicitly instructed rule-breaking, to mimic deception outside of the laboratory.

Cognitive and neuroscientific researchers have begun to explore the role of voluntary dishonest behavior (Blakemore, Winston, & Frith, 2004; Kozel et al., 2005; Mohamed et al., 2006; Sip et al., 2008). More recently, the field of *detecting* deception has also started to focus on more ecologically valid paradigms to mimic deceit and investigate its role in test sensitivity. For instance, Nahari, Breska, Elber, Klein Selle, and Ben-Shakhar (2017) let participants 'choose' to either enact a mock crime or an innocent computerized task. The study revealed a similar CIT detection efficiency, based on psychophysiological and RT measures, for those who choose to commit the mock crime as for participants who were explicitly ordered to commit the mock crime.

Yet in forced-choice-paradigms, even 'deceptive' participants are not guilty rule-breakers, but in fact obediently complying with the experimenters' instructions. It is pointed out that deception, in all its complexity, can only be fully investigated when the decision

to deceive is based on the participants' own initiative. The current study attempts to address this limitation by investigating whether the act of deliberate versus obedient cheating in this new paradigm influences the sensitivity of deception detection techniques based on RT measures.

#### *1.4. Current study*

The vast majority of deception studies have relied upon instructed deception. In an attempt to enhance external validity while maintaining experimental control, self-initiated deception in the current study was defined as cheating on a trivia quiz. Self-initiated cheaters were compared with instructed cheaters and non-cheaters (i.e., fair players).

By borrowing successful cheating paradigms from the field of social psychology and behavioral economics (see DeAndrea, Carpenter, Shulman, & Levine, 2009; Domnich et al., 2015; Halevy, Shalvi, & Verschuere, 2014; Nagin & Pogarsky, 2003), participants in the current experiment were provided with a monetary incentive for high accuracy performance on a trivia quiz. Unbeknownst to the participants, the quiz was constructed in a way that it would be almost impossible to earn the bonus without looking up the correct answers on the last two questions.

Participants were randomly allocated to either a condition where they were instructed to cheat on the quiz (mimicking the typical laboratory set-up) or to a condition in which they were provided with the opportunity to cheat, yet without explicit instructions to do so. In this manner, deceptive behavior was completely self-initiated and involved the crucial deliberate intent to deceive. Moreover, contrary to most laboratory research, participants in the current study were not aware of the fact that they would partake in a memory detection test during the experiment. This way, critical items were incidentally acquired and thereby both encoding and retention resembled realistic settings (see Meixner & Rosenfeld, 2014).

By mirroring the conditions that would be present in a real-world environment, individual differences in the decision to cheat were examined as a secondary aim of this study. According to Lee and Ashton (2012), individual differences in honesty and morality are a core component of personality, which they introduced as an additional factor to the Big Five personality inventory. This specific new factor, called Honesty-Humility (HH) in the HEXACO Personality Inventory (HEXACO-PI-R; Lee & Ashton, 2004), taps into differences in the willingness to manipulate others for personal gain, the temptation to break rules and interest in social status. Across six studies, a consistent negative correlation was found between HH scores and the likelihood of cheating (Hilbig & Zettler, 2015). In the current study, HH scores of self-initiated cheaters and fair players are compared as to further explore differences between participants who willingly cheat or not. It is expected that self-initiated cheaters report to be more inclined to break rules for financial gain compared to fair players, reflected by lower HH scores in the HEXACO-PI-R. Furthermore, these two groups are compared to the instructed cheaters, who form the baseline condition for this particular measure. This allows us to clarify whether individual

differences in cheating behavior are driven by personality attributes of the cheaters (as often assumed) or of the fair players.

## 2. Method

The study was approved by the ethical committee of the Department of Psychology of the University of Amsterdam (2016-CP-7217). The tasks scripts and data are available on <https://osf.io/t9y7n>.

### 2.1. Participants

The sample consisted of 433 students (79.4% female) from the University of Amsterdam. Their average age was 20.41 years old ( $SD$  age = 2.41, range from 17 to 44). Participants were recruited through a university portal and received course credits as compensation. Participants were mainly students pursuing a degree in psychology (52%), communication science (21%) or psychobiology (21%). All participants provided consent before taking part in the study.

Participants were randomly allocated to the instructed cheating versus the possibility to cheat condition with a 1:3 ratio, as the latter condition was subsequently split in self-initiated cheaters versus fair players.

One of the two cheating-evoking questions of the trivia quiz asked about the author of the novel *Wishful Drinking*. However, halfway during data collection the author passed away, which evoked multiple news items occasionally mentioning her novel. Because of possible familiarity with the probe, we excluded the data of 44 participants that were tested in the week after Carrie Fisher's illness and death. For all subsequent participants, this trivia question was replaced with an equally difficult question.

#### 2.1.1. Instructed cheaters, self-initiated cheaters, and fair players

Based upon the instructions and their performance on the cheating-evoking trivia quiz questions, participants were labeled instructed cheaters (i.e., those given instructions to cheat), self-initiated cheaters (i.e., those given opportunity to cheat and answered the two cheating-evoking questions correctly) or fair players (i.e., those given opportunity to cheat and answered neither of the two cheating-evoking questions correctly). Twenty-two individuals answered only one of the two cheating-evoking questions correctly and were excluded from further analysis due to an uncertain ground truth criterion regarding their cheating behavior.

There were 130 participants in the instructed cheating condition (17.7% male,  $M$  age = 20.57,  $SD$  age = 2.83), who completed the HEXACO and trivia quiz. CIT data were missing from three participants and 20 participants were excluded due to low target accuracy<sup>1</sup> (i.e., an error rate of 50% or more on target items, see Kleinberg & Verschuere, 2015). The final sample for CIT analysis consisted therefore of 107 instructed cheaters (20.6% male,  $M$  age = 20.72,  $SD$  age = 3.03).

There were 259 participants in the possibility to cheat condition (22.8% male,  $M$  age = 20.37,  $SD$  age = 1.95). One hundred seventy-five participants answered both cheating-evoking questions of the trivia quiz correctly and were named the self-initiated cheaters (23.4% male;  $M$  age = 20.04,  $SD$  age = 1.73). All completed the HEXACO and trivia quiz. CIT data were missing from eight participants and 25 participants were excluded due to low target accuracy. The final sample for CIT analysis consisted therefore of 142 self-initiated cheaters (22.5% male,  $M$  age = 19.94,  $SD$  age = 1.65).

Sixty-two participants (21% male;  $M$  age = 20.59,  $SD$  age = 2.32) did not know the correct answers to the difficult quiz-questions and were therefore labeled fair players. One participant responded incorrectly to two out of three validation questions in the personality measure and was therefore excluded from HEXACO analysis. CIT data were missing from four participants and 13 participants were excluded due to low target accuracy. The final sample for CIT analysis therefore consisted of 45 fair players (17.8% male,  $M$  age = 20.55,  $SD$  age = 1.89).

There were no significant differences in gender between the three groups,  $X^2(2) = 1.48$ ,  $p = .477$ ,  $\phi_c = 0.06$ . Of all 237 participants in the possibility to cheat condition (partial cheaters excluded), 41 out of 54 male participants cheated on the trivia quiz (75.9%). For female participants, 134 out of 183 (73.2%) cheated.

The three groups differed significantly in age, Welch's  $F(2, 144.29) = 4.80$ ,  $p = .010$ , est.  $\omega^2 = 0.21$ . A Games-Howell post hoc test revealed that self-initiated cheaters were significantly younger than fair players ( $p = .018$ ,  $d = 0.45$ ). Both groups did not differ significantly in age from instructed cheaters ( $p = .147$ ,  $d = 0.23$  and  $p = .591$ ,  $d = 0.14$ , respectively). Note that data of five participants were missing from this analysis. As the absolute age difference is minute and the potential impact within this range is very limited, age was not included as a covariate in the reported analyses.

## 2.2. Material

### 2.2.1. HEXACO

Personality traits were assessed with the 60-item Dutch version of the HEXACO Personality Inventory Revised (Ashton & Lee, 2009). This inventory measures the six major dimensions of personality: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to experience. Participants had to indicate to what extent they agreed or disagreed with each statement on a 5-point Likert scale ranging from 1 (*Completely disagree*) to 5 (*Completely agree*). Each scale showed a good internal consistency using Cronbach's alpha (Cronbach, 1951): Honesty-Humility ( $\alpha = 0.71$ ), Emotionality ( $\alpha = 0.79$ ), Extraversion ( $\alpha = 0.82$ ), Agreeableness ( $\alpha = 0.77$ ), Conscientiousness ( $\alpha = 0.80$ ), and Openness to experience ( $\alpha = 0.76$ ).

### 2.2.2. Trivia quiz

Participants engaged in a trivia quiz with 10 open questions. Participants were not aware of the fact that the quiz was constructed in a way that it would be almost impossible to earn the bonus without looking up the correct answers to the last two questions

online. The chosen eight easy questions were correctly solved by approximately 90% of the participants in a pilot study. The two questions that were each correctly solved by no more than 6% of the pilot's participants were classified as cheating-evoking.<sup>2</sup> Therefore, the probability of answering all questions correctly without cheating would be statistically highly unlikely. As a result, participants claiming the trivia bonus for answering all 10 questions correctly, including the two cheating-evoking questions ('Who coined the term *dinosaur*?' as well as 'Who wrote the autobiographical book *Prairie Tale: A Memoir*?' as a replacement for 'Who wrote the autobiographical book *Wishful Drinking*?'), were assumed to be cheaters. All items were pretested to ensure that the correct answers could be found within the first three search results on Google (see Domnich et al., 2015).

### 2.2.3. CIT

The Concealed Information Test (CIT) is a method designed to measure concealed knowledge using RTs. The objective of the CIT is to verify whether the suspect possesses crime-related information that only the perpetrator would be aware of. The method requires that the examiner identifies a number of established facts from the investigation which only the true culprit will be able to recognize. Then, the examiner presents various crime-related details (*probes*) embedded in a set of foil items (*irrelevants*) while measuring the suspects' behavioral responses. By choosing irrelevant items carefully, in a way that all options would seem equally plausible to unknowledgeable individuals, all items should trigger a similar response pattern. By denying the crime-relevant items (i.e., responding "no" to probes instead truthfully admitting recognition), guilty suspects actively conceal information.

Whereas the trivia quiz manipulated the conscious act of cheating for an incentive versus fair playing, during the CIT all participants were explicitly instructed to conceal their knowledge of the correct items. Participants are required to deny knowledge both for trials containing critical items (i.e., respond "no", hence lying) as well as irrelevant items (i.e., respond "no", hence telling the truth). Lastly, *target* items are added to ensure that examinees pay attention to all items. These items have to be answered "yes" and are learned just before commencing the CIT.

### 2.2.4. Follow-up questionnaire

Motivational states were reported in a questionnaire involving five questions that participants had to rate on a 5-point Likert scale. This questionnaire measured how well participants were able to focus on the screen during the CIT, how involved they were in the study, how well their memory was for the items of the trivia quiz and the learned target items, and how much they tried to avoid detection and appear innocent on the CIT.

### 2.2.5. Recall and recognition

Memory for the items of the trivia quiz was assessed with a free recall followed by a recognition test. Participants first had to recall the correct answer to the two easy as well as the two cheating-evoking questions from the trivia quiz that were subsequently used in



the CIT. Afterwards, participants had to pick the correct option when presented with the probe and the four irrelevant options.

For recall, answers to the questions were each coded as either correct (1) or incorrect (0), leading to a score between 0–2 per item type (i.e., easy and cheating-evoking questions). Using arbitrary criteria, answers were coded correct if participants recalled both the first and last name correctly (e.g., Carrie Fisher as the author of the novel *Wishful Drinking*) or when participants only recalled the last name correctly (e.g., Fisher). When an incorrect first name was entered in combination with a correct last name, or if only the first name was mentioned, the recall was coded as incorrect. For recognition, items were scored as either correct (1) or incorrect (0), leading to a score between 0 and 2 per item type (i.e., easy and cheating-evoking questions).

### 2.3. Procedure

Participants signed up for a study concerning the relationship between mood and its impact on task performance, since announcing the true purpose of the study could distort the goal to investigate self-initiated deceptive behavior. Participants completed the study online at their own time and on their own computer with keyboard. Approximate total participation time was 60 min.

The first part of the experiment was hosted on the survey website Qualtrics. After reading the information brochure and signing the informed consent, participants started out with the 60-item Dutch version of the HEXACO Personality Inventory Revised (De Vries, Lee, & Ashton, 2008). To control for serious participation, three attentiveness checks were inserted (e.g., this is a control question, please indicate ‘I completely agree’). Moreover, participants could not continue to the next page within 30 s to ensure proper inspection of the material.

Then participants engaged in two seemingly relevant problem solving tasks inserted to disguise the true purpose of the study. Participants were told that they must complete these tasks alone within the timeframe of 5 min each. This was followed by a trivia quiz with 10 open (eight easy and two difficult, cheating evoking) questions. It was again made clear to the participants that they should complete the tasks alone within the timeframe of 10 min, thereby indirectly communicating that it was not allowed to use Google. Moreover, an incentive in course credits was promised when all 10 questions would be answered correctly.

Various tactics were inserted to increase the chance of cheating. To ensure a sense of privacy and anonymity, demographics were not inquired until after the experiment. Since a clear and fast reward has been shown to increase dishonest behavior (Gino & Pierce, 2009), an incentive was offered to participants who answered all trivia questions correctly. Also, since the odds of cheating are higher at the end of a series than earlier, the cheating-evoking questions were presented as the last two questions of the quiz. The temptation to cheat is maximized when people are faced with the final opportunity to gain a reward (Effron, Bryan, & Murnighan, 2015). Lastly, participants could not continue to the next assignment within the 10-min time frame. Pilot studies showed that this is

enough time to finish the first eight easy questions (roughly two min), while leaving ample time to retrieve the correct answers online. Moreover, by setting the page timer to 10 min, sufficient time was created for participants to take an active decision of whether to cheat or not. Perhaps this also led to some frustration, thereby possibly providing entitlement and self-justification for cheating behavior (see also Mazar, Amir, & Ariely, 2008).

For pilot purposes, participants were asked to report as specifically as possible how they knew the answers to four randomly chosen questions from the trivia quiz. In reality, these were always the same two easy and two cheating-evoking questions that were later used in the CIT. This gave participants the opportunity to either confess to cheating, deny their cheating, or repeat that they did not know the answer. Participants also completed the 20-item Dutch version of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) following the HEXACO, the quiz, and the opportunity to confess. Besides serving for pilot purposes, this additional mood questionnaire facilitated the cover story, concealing the true purpose of the study.

Then, participants continued to the second part of the experiment, starting with the memory detection test, programmed in Inquisit 4.0 by Millisecond Software. Participants were told to hide their knowledge of the correct answers in the trivia quiz they completed previously. Upon successful concealment,<sup>3</sup> an additional incentive in course credits would be awarded. The CIT included answers to two of the easy questions as to measure a baseline-response to known items, as well as the answers to the two cheating-evoking questions that participants might have cheated on. Per question, the CIT included the correct answer, four incorrect answers serving as irrelevant options, and a target item. Hence, if Rome was the probe stimulus, the target was Madrid, and the irrelevant stimuli were Paris, Vienna, London, and Berlin.

Participants had to respond to the question “*Is this the correct answer?*” by pressing either the left (A-key) button for YES or the right (L-key) button for NO on their keyboard. The question and the response keys remained on the screen during the entire test as a reminder. Participants were instructed to respond with YES only to the target items and NO to all other stimuli (i.e., both the correct answers of the trivia quiz and the irrelevant options). Each trial consisted of one answer (e.g., Berlin) being displayed as a word in the middle of the screen for exactly 1,500 ms. If the participant did not give a response within the maximum response deadline of 800 ms, the message TOO SLOW appeared in red color above the stimulus for 200 ms. If the participant’s response was incorrect, that is responding with NO for target items or with YES to probe or irrelevant items, the word WRONG appeared in red color below the stimulus for 200 ms. Response latency was measured from the onset of the stimulus on the screen until one of the response keys was pressed. After key-press or after the 1,500 ms presentation time, the next stimulus appeared on the screen with an inter-stimulus interval (ISI) of either 250, 500, or 750 ms to prevent response preparation.

In order to ensure proper understanding of the task and instructions, each participant had to pass through a stepwise practice procedure that allowed participants to become used to the speed and requirements of the task. Each of the three practice phases of the

memory detection test consisted of 24 trials. In the first practice phase, participants could pace the speed of the trial sequence themselves, so that a new stimulus only appeared after a key press. Feedback was given upon an erroneous response, but the TOO SLOW message was not presented. Participants could proceed to the next phase when their target accuracy was at least 75%; otherwise the first practice phase was repeated until this requirement was met. In the second phase the 1,500 ms stimulus presentation was added, so that the next trial would automatically appear upon key press or after 1,500 ms. Again, feedback was given upon an erroneous response, but the TOO SLOW message was never presented. Participants could only proceed to the next phase when their target accuracy was at least 50% and as an additional requirement, when their mean response latency was below 800 ms, otherwise this practice phase was repeated until their performance was satisfactory. The last practice phase was identical to the full test, including the WRONG and TOO SLOW feedback. Participants could proceed to the actual test only when their target accuracy was at least 50% and when their mean response latency was below 800 ms.

In the actual CIT, all words were presented in a 1:1:4 ratio; that is, of the total 480 trials in the test, 80 were probe stimuli, 80 were target stimuli, and 320 were irrelevant stimuli, so that each stimulus was displayed exactly 20 times. The sequence of the stimuli was randomized, as well as the question types (i.e., easy and cheating-evoking) and the ISI.

After completing the CIT, participants were presented with a questionnaire designed to assess their attention to the tasks, involvement in the experiment, memory for the stimuli, and their motivation to avoid detection in the CIT on a 5-point Likert scale (ranging from 1 = *not at all* to 5 = *very much so*). Then, participants were told that the experiment was finished and they did not have to hide any information anymore. They had to give honest answers to the four CIT questions (i.e., the two easy questions as well as the two cheating-evoking questions) in a free-recall and subsequently in a multiple-choice recognition test. Only thereafter demographic information was asked, including gender, age, and field of study. Lastly, participants were debriefed and compensated for participation.

### 3. Results

#### 3.1. CIT

Trials to which no response was recorded (i.e., RTs larger than 1,500 ms) were excluded from all subsequent analyses. Moreover, trials with an incorrect response (i.e., pressing NO for target items and/or YES for probe and irrelevant items) as well as trials with a RT below 150 ms and above 800 ms (see also Verschuere, Crombez, Degrootte, & Rosseel, 2010) were excluded from response latency analysis. Of all 294 participants who completed the CIT, three participants did not give the correct answer to one of the easy questions, and CIT trials addressing those questions were excluded from analysis.

CIT trials addressing the cheating-evoking questions for fair players who admitted to have searched for the answer online without filling it in on the trivia quiz were excluded from the analysis. On average, 445 trials (92.7%) per participant were included in the analyses (range: 61.5%–98%). All analyses used an alpha level of 0.05. Effect sizes for the ANOVA are reported using Cohen's  $f$ . For follow-up contrasts Cohen's  $d$  is used.<sup>4</sup> Cohen's  $d$  for within-subject and between-subject comparisons is annotated as  $d_{\text{within}}$  and  $d_{\text{between}}$ . As a rule of thumb, Cohen (1988) proposed 0.20, 0.50 and 0.80 as thresholds for "small," "moderate," and "large" effects, respectively, for  $d$  values and 0.10, 0.25, and 0.40 as thresholds for "small," "moderate," and "large" effects, respectively, for  $f$  values. Moreover, JZS Bayes factors were computed, using JASP software version 0.8.4, which are numerical values quantifying the odds ratio between the null and the alternative hypothesis given the data. A default JZS prior with scaling factor  $r = 0.707$  was used for the alternative hypothesis (see Rouder, Speckman, Sun, Morey, & Iverson, 2009). The JZS Bayes factors are reported as either in favor of the null or the alternative hypothesis. Using Jeffreys's (1961) criteria, a Bayes factor of three or more is taken as substantial evidence for the respective hypothesis. Lastly, the area under the Receiver Operating Characteristic (ROC) curve was calculated. This statistic describes the detection efficiency of the CIT in differentiating between knowledgeable and unknowledgeable individuals, computed across all possible cut-off points on the detection score. The ROC area varies between 0 and 1, with a chance level of 0.5 (for a more detailed description, see Liebllich, Kugelmass, & Ben-Shakhar, 1970).

### 3.1.1. RTs

In the main analysis, a 3 (Condition: self-initiated cheater vs. instructed cheater vs. fair player, between-participants) by 2 (Stimulus: probe vs. irrelevant, within-participants) by 2 (Question: easy vs. cheating, within-participants) mixed ANOVA was conducted on reaction times in milliseconds. Although RTs were the prime outcome measure as they are typically more valid than error rates (Kleinberg & Verschuere, 2015), the same ANOVA was conducted and reported for error rates.

The mixed ANOVA revealed a significant main effects of Stimulus,  $F(1, 288) = 188.59$ ,  $p < .001$ ,  $f = 0.81$  (i.e., longer RTs to probes than to irrelevants), and Question,  $F(1, 288) = 136.48$ ,  $p < .001$ ,  $f = 0.69$  (i.e., longer RTs to the cheating-evoking questions than to the easy questions). Significant interactions were revealed between Condition and Stimulus,  $F(2, 288) = 5.48$ ,  $p = .005$ ,  $f = 0.20$  (i.e., larger probe-irrelevant difference for self-initiated and instructed cheaters than for fair players), and between Condition and Question,  $F(2, 288) = 3.33$ ,  $p = .037$ ,  $f = 0.15$  (i.e., greater difference in RTs between the cheating-evoking and easy questions for self-initiated and instructed cheaters compared to fair players). These effects are collapsed under a statistically significant three-way interaction of Condition by Stimulus by Question,  $F(2, 288) = 7.38$ ,  $p = .001$ ,  $f = 0.23$ . Table 1 shows the mean RTs for each cell of the design.

To narrow down the three-way interaction, a 3 (Condition: self-initiated cheater vs. instructed cheater vs. fair player, between-participants) by 2 (Stimulus: probe vs.



irrelevant, within-participants) mixed ANOVA was conducted for the easy and the cheating-evoking questions separately.

For the easy questions, the ANOVA revealed a significant main effect of Stimulus,  $F(1, 291) = 94.18$ ,  $p < .001$ ,  $f = 0.57$ , and no significant interaction between Condition and Stimulus  $F(2, 291) = 0.14$ ,  $p = .867$ ,  $f = 0.03$ . This indicates that there was a clear CIT-effect across the conditions and since all participants knew the correct answer, no differences emerged between the three conditions.

For the cheating-evoking questions, the ANOVA revealed a significant main effect of Stimulus,  $F(1, 288) = 89.31$ ,  $p < .001$ ,  $f = 0.56$ , and a significant interaction between Condition and Stimulus  $F(1, 288) = 13.33$ ,  $p < .001$ ,  $f = 0.30$ . This indicates that the CIT-effect differed according to the condition. Planned contrasts in a follow-up one-way ANOVA on RT difference scores (i.e.,  $RT_{\text{probe}} - RT_{\text{irrelevant}}$ ) revealed a significant difference between the two conditions in which participants knew the correct answer to the cheating-evoking questions (i.e., self-initiated and instructed cheaters) versus the fair players, who did not know the correct answers,  $t(288) = 5.08$ ,  $p < .001$ ,  $d_{\text{between}} = 0.83$ ,  $BF = 19,512.42$  in favor of the alternative hypothesis. There was no significant difference in the RTs between the self-initiated and instructed cheaters,  $t(288) = -1.20$ ,  $p = .232$ ,  $d_{\text{between}} = 0.15$ ,  $BF = 3.62$  in favor of the null hypothesis.

### 3.1.2. ROC

In order to analyze detection efficiency of classifying individuals as knowledgeable (i.e., cheaters) versus unknowledgeable (i.e., fair players), we compared the distribution of the within-person detection scores for both self-initiated cheaters and instructed cheaters with the detection score distribution of fair players. For both of these comparisons, we computed the ROC area as well as their respective 95% confidence intervals. Analysis reveals that detection efficiency was significantly larger than a chance area of 0.50 for both self-initiated and instructed cheaters with ROC areas of 0.73 [0.64; 0.81] and 0.78 [0.70; 0.86], respectively.

### 3.1.3. Error rates

The 3 (Condition: self-initiated cheater vs. instructed cheater vs. fair player, between-participants) by 2 (Stimulus: probe vs. irrelevant, within-participants) by 2 (Question: easy vs. cheating, within-participants) mixed ANOVA on error rates revealed a significant main effect of Stimulus,  $F(1, 291) = 148.99$ ,  $p < .001$ ,  $f = 0.72$  (i.e., more errors to probes than to irrelevants), that was qualified by the Condition  $\times$  Question,  $F(2, 291) = 3.59$ ,  $p = .029$ ,  $f = 0.16$ , and Stimulus  $\times$  Question,  $F(1, 291) = 4.39$ ,  $p = .037$ ,  $f = 0.12$  interactions. These effects are collapsed under the three-way significant interaction of Condition  $\times$  Stimulus  $\times$  Question,  $F(2, 291) = 6.71$ ,  $p = .001$ ,  $f = 0.21$ . Table 1 shows the mean errors for each cell of the design.

To narrow down the three-way interaction, a 3 (Condition: self-initiated cheater vs. instructed cheater vs. fair player, between-participants) by 2 (Stimulus: probe vs. irrelevant, within-participants) mixed ANOVA was conducted for the easy and the cheating-evoking questions separately.

For the easy questions, the ANOVA revealed a significant main effect of Stimulus,  $F(1, 291) = 102.25$ ,  $p < .001$ ,  $f = 0.59$ , and no significant interaction between Condition and Stimulus  $F(2, 291) = 1.88$ ,  $p = .155$ ,  $f = 0.11$ . This indicates that there was a clear CIT-effect across the conditions and since all participants knew the correct answer, no differences emerged between the three conditions.

For the cheating-evoking questions, the ANOVA revealed a significant main effect of Stimulus,  $F(1, 291) = 47.44$ ,  $p < .001$ ,  $f = 0.40$ , and a significant interaction between Condition and Stimulus  $F(2, 291) = 5.55$ ,  $p = .004$ ,  $f = 0.20$ . Planned contrasts in a follow-up one-way ANOVA conducted on the error rate difference scores (i.e., error rate<sub>probe</sub> – error rate<sub>irrelevant</sub>) revealed a significant difference between the two conditions in which participants knew the correct answer to the cheating-evoking questions (i.e., self-initiated and instructed cheaters) versus the fair players, who did not know the correct answers,  $t(291) = 3.33$ ,  $p = .001$ ,  $d_{\text{between}} = 0.54$ ,  $BF = 27.32$  in favor of the alternative hypothesis. There was no significant difference in the error rates difference scores between self-initiated and instructed cheaters,  $t(291) = -0.04$ ,  $p = .972$ ,  $d_{\text{between}} = 0.00$ ,  $BF = 7.13$  in favor of the null hypothesis.

### 3.2. Hexaco

A one-way ANOVA was conducted on the data of the 366 participants who successfully completed the HEXACO personality measure, to determine whether reported Honesty-Humility differed between the three experimental conditions. Analysis revealed no statistically significant difference between the three conditions on Honesty-Humility,  $F(2) = 1.95$ ,  $p = .144$ ,  $f = 0.11$ .

For completion sake, we contrasted the groups on Honesty-Humility. The planned contrast did not show a statistically significant difference between self-initiated cheaters ( $M = 3.18$ ,  $SD = 0.54$ ) and fair players ( $M = 3.30$ ,  $SD = 0.56$ ,  $t(363) = -1.51$ ,  $p = .133$ ,  $d_{\text{between}} = 0.22$ ,  $BF = 2.15$  in favor of the null hypothesis). Post hoc comparisons with Bonferroni correction revealed that self-initiated cheaters as well as fair players did not differ from the control group (i.e., instructed cheaters;  $M = 3.14$ ,  $SD = 0.53$ ,  $p = 1.00$ ,  $d_{\text{between}} = 0.07$ , and  $p = .149$ ,  $d_{\text{between}} = 0.30$ , respectively).

Also for sake of completion, additional ANOVAs showed no significant effects for the remaining personality factors, Emotionality,  $F(2) = 0.13$ ,  $p = .881$ ,  $f = 0.03$ , Agreeableness,  $F(2) = 0.28$ ,  $p = .760$ ,  $f = 0.04$ , Conscientiousness,  $F(2) = 0.38$ ,  $p = .684$ ,  $f = 0.04$ , or Openness to Experience,  $F(2) = 1.52$ ,  $p = .221$ ,  $f = 0.09$ . There was a statistically significant difference between the three conditions on Extraversion, Welch's  $F(2, 147.10) = 3.59$ ,  $p = .030$ , est.  $\omega^2 = 0.01$ . A Games-Howell post hoc test that contrasted self-initiated cheaters ( $M = 3.62$ ,  $SD = 0.54$ ) with fair players ( $M = 3.37$ ,  $SD = 0.78$ ) revealed a marginally significant effect of Extraversion ( $p = .053$ ,  $d_{\text{between}} = 0.41$ ). No significant differences in levels of Extraversion emerged between the self-initiated cheaters and fair players with the control group of instructed cheaters ( $M = 3.51$ ,  $SD = 0.54$ ,  $p = .169$ ,  $d_{\text{between}} = 0.20$  and  $p = .402$ ,  $d_{\text{between}} = 0.22$ , respectively). Table 2 shows the descriptive statistics and internal consistency.

Table 2

Mean scores (*SDs* in Parentheses) on the factors of 60-item Dutch version of the HEXACO-PI-R on a 5-point Likert scale

	Self-Initiated Cheaters ( <i>n</i> = 175)	Fair Players ( <i>n</i> = 61)	Instructed Cheaters ( <i>n</i> = 130)	<i>p</i> -value	Effect Size	Internal Consistency
Honesty-humility	3.18 (0.54)	3.30 (0.56)	3.14 (0.53)	<i>p</i> = .144	<i>f</i> = 0.11	$\alpha$ = 0.71
Emotionality	3.24 (0.62)	3.29 (0.52)	3.26 (0.66)	<i>p</i> = .881	<i>f</i> = 0.03	$\alpha$ = 0.79
Extraversion	3.62 (0.54)	3.37 (0.78)	3.51 (0.54)	<i>p</i> = .030	$\omega^2$ = 0.01	$\alpha$ = 0.82
Agreeableness	3.13 (0.58)	3.10 (0.60)	3.08 (0.59)	<i>p</i> = .760	<i>f</i> = 0.04	$\alpha$ = 0.77
Conscientiousness	3.47 (0.65)	3.39 (0.62)	3.47 (0.61)	<i>p</i> = .684	<i>f</i> = 0.04	$\alpha$ = 0.80
Openness to experience	3.33 (0.61)	3.49 (0.63)	3.41 (0.66)	<i>p</i> = .221	<i>f</i> = 0.09	$\alpha$ = 0.76

Table 3

Mean scores (*SDs* in parentheses) on the follow-up questionnaire on a 5-point Likert scale

	Self-Initiated Cheaters ( <i>n</i> = 141)	Fair Players ( <i>n</i> = 45)	Instructed Cheaters ( <i>n</i> = 107)	<i>p</i> -value	Effect Size
Reported focus on the computer screen	3.48 (1.17)	3.58 (1.27)	3.66 (1.17)	<i>p</i> = .462	<i>f</i> = 0.07
Reported involvement in the experiment	4.04 (0.89)	4.00 (1.00)	4.04 (0.91)	<i>p</i> = .970	<i>f</i> = 0.00
Reported memory for the answers of the quiz	3.96 (0.89)	3.56 (1.01)	3.97 (0.81)	<i>p</i> = .043	$\omega^2$ = 0.02
Reported memory for the target items	4.09 (0.85)	4.16 (0.90)	4.15 (0.87)	<i>p</i> = .806	<i>f</i> = 0.03
Reported effort to hide knowledge	4.42 (0.71)	4.29 (1.01)	4.32 (0.90)	<i>p</i> = .649	<i>f</i> = 0.06

### 3.3. Follow-up questionnaire

Of all 294 participants who successfully completed the CIT, one participant did not complete the follow-up questionnaire and was therefore excluded from this analysis. Table 3 shows the mean scores for each cell of the design.

The three conditions did not differ in their reported focus,  $F(2) = 0.77$ ,  $p = .462$ ,  $f = 0.07$ , involvement,  $F(2) = 0.03$ ,  $p = .970$ ,  $f = 0.00$ , memory for the target items,  $F(2) = 0.22$ ,  $p = .806$ ,  $f = 0.03$ , or motivation to avoid detection,  $F(2) = 0.649$ ,  $p = .523$ ,  $f = 0.06$ . Analysis revealed a statistically significant difference between the three conditions on their reported memory for the answers of the trivia quiz, Welch's  $F(2, 115.59) = 3.49$ ,  $p = .043$ , est.  $\omega^2 = 0.02$ .

A Games-Howell post hoc test showed that the fair players ( $M = 3.56$ ,  $SD = 1.01$ ) differed significantly from both instructed ( $M = 3.97$ ,  $SD = 0.81$ ,  $p = .044$ ,  $d_{\text{between}} = 0.47$ ) and self-initiated cheaters ( $M = 3.96$ ,  $SD = 0.89$ ,  $p = .051$ ,  $d_{\text{between}} = 0.43$ ) on reported memory for the answers of the quiz. The latter two conditions did not differ significantly ( $p = .990$ ,  $d_{\text{between}} = 0.01$ ).



### 3.4. Recall and recognition

Of all 294 participants who successfully completed the CIT, one participant did not complete the free recall and recognition tasks that assessed memory for the four critical items of the trivia quiz.

#### 3.4.1. Recall

A one-way ANOVA for the easy questions revealed a clear ceiling effect for self-initiated cheaters ( $M = 1.99$ ,  $SD = 0.08$ ), instructed cheaters ( $M = 1.96$ ,  $SD = 0.23$ ), and fair players ( $M = 1.89$ ,  $SD = 0.38$ ), with no significant difference in recall rate between the conditions, *Welch's F* (2, 88.71) = 2.35,  $p = .101$ , est.  $\omega^2 = 0.01$ . A second one-way ANOVA for the cheating-evoking questions revealed a significant difference between self-initiated cheaters ( $M = 1.64$ ,  $SD = 0.69$ ), instructed cheaters ( $M = 1.55$ ,  $SD = 0.73$ ), and fair players ( $M = 0.16$ ,  $SD = 0.47$ ), *Welch's F* (2, 142.53) = 149.11,  $p < .001$ , est.  $\omega^2 = 0.50$ . A Games-Howell post hoc test showed that the fair players scored significantly lower than both instructed ( $p < .001$ ,  $d_{\text{between}} = 2.08$ ) and self-initiated cheaters ( $p < .001$ ,  $d_{\text{between}} = 2.29$ ). The latter two conditions did not differ significantly ( $p = .609$ ,  $d_{\text{between}} = 0.13$ ).

#### 3.4.2. Recognition

A one-way ANOVA for the easy questions revealed a ceiling effect of recognition for self-initiated cheaters ( $M = 1.99$ ,  $SD = 0.08$ ), instructed cheaters ( $M = 1.98$ ,  $SD = 0.14$ ), and fair players ( $M = 1.98$ ,  $SD = 0.15$ ), with no significant difference in recognition rate between the conditions,  $F(2) = 0.45$ ,  $p = .641$ ,  $f = 0.05$ . A second one-way ANOVA for the cheating-evoking questions revealed a significant difference between self-initiated cheaters ( $M = 1.98$ ,  $SD = 0.14$ ), instructed cheaters ( $M = 1.98$ ,  $SD = 0.14$ ) and fair players ( $M = 0.78$ ,  $SD = 0.70$ ), *Welch's F* (2, 99.82) = 64.80,  $p < .001$ , est.  $\omega^2 = 0.30$ . A Games-Howell post hoc test showed that the fair players scored significantly lower than both instructed ( $p < .001$ ,  $d_{\text{between}} = 3.01$ ) and self-initiated cheaters ( $p < .001$ ,  $d_{\text{between}} = 3.29$ ). The latter two conditions did not differ significantly ( $p = .989$ ,  $d_{\text{between}} = 0.00$ ).

## 4. Discussion

To examine the external validity of detection of deception research, the current study explored whether the cognitive signature of lying differs for instructed versus self-initiated cheating. Using a well-established paradigm (CIT; Lykken, 1959; Verschuere, Ben-Shakhar, & Meijer, 2011) and measure (RTs; Suchotzki et al., 2017), we instructed some participants to cheat on a trivia quiz, while providing others with an opportunity and incentive to cheat without an explicit instruction. Results of 294 memory detection tests replicated the typically observed response pattern (i.e., increased response latency for relevant items), with no differences between instructed and self-initiated cheaters. This result indicates that the cognitive signature of lying extends to self-initiated cheating,

thereby strengthening the external validity of CIT studies using RTs. Additionally, the CIT adequately discriminated between cheating participants and fair players.

#### 4.1. *The cognitive signature of lying*

From a cognitive perspective, lying is regarded as a cognitive process that imposes a greater burden on our executive functions than truth telling, which is viewed as the default in human communication. A growing literature, using a diverse array of paradigms and measures, provides increasing support for the idea that lying is typically more demanding than truth telling (Caso et al., 2005; Debey, Ridderinkhof, De Houwer, De Schryver, & Verschuere, 2015; Debey, Verschuere, & Crombez, 2012; Duran et al., 2010; Vrij et al., 1996; Walczyk et al., 2003). In the current paradigm this response conflict is reflected by the prolonged RT when deliberately denying knowledge compared to responding to neutral irrelevant options. Using this contrast between probes and irrelevant items, the CIT distinguished cheaters from fair players ( $d = 0.83$ ).

While various approaches using the cognitive load theory have emerged, the specific underlying mechanisms that generate the difference between liars and truth-tellers remains debated. Researchers have summoned to investigate how various executive functions such as working memory and response inhibition exactly contribute to this cognitive signature of lying (see Blandón-Gitlin, Fenn, Masip, & Yoo, 2014; Gombos, 2006) and whether boundary conditions apply. By investigating the influence of self-initiated (as opposed to laboratory-induced) behavior, the current findings shed a light on the role of intent on the cognitive processes underlying deception.

Recent insights suggest that Response Inhibition (RI; Verschuere, Crombez, Koster, Van Bockstaele, & De Clercq, 2007) plays a critical role in obtaining the cognitive signature of lying, in particular for the RT-CIT (Suchotzki et al., 2017). The RT-CIT has been found to be influenced by neither item saliency (Klein Selle et al., 2017) nor by increased motivation to avoid detection (Kleinberg & Verschuere, 2016). This could explain why no differences in detection accuracy were found between the participants who were instructed to cheat on the quiz (mimicking the typical laboratory set-up) and the participants who cheated on the quiz on their own initiative ( $d = 0.15$ , as well as substantial evidence for the null hypothesis;  $BF = 3.62$  in favor of the null hypothesis).

The data corroborate the hypothesis of the cognitive impact on lying, with the addition that it is not restricted to explicitly instructed cheating, but can also be observed for intentional, self-initiated cheating.

#### 4.2. *Who cheats?*

Literature in clinical psychology has focused on personality traits that causes some individuals to cheat with a higher likelihood than others. For instance, psychopathy (Nathanson, Paulhus, & Williams, 2006) has been linked to increased likelihood of cheating. In fact, individual propensity to cheat has been deemed so important that Honesty-Humility was added as an additional factor to the Big Five in the HEXACO model

of personality (Hilbig & Zettler, 2015). Social psychology, on the other hand, has focused on state factors that prompt individuals to cheat, depending on the situational context. For example, cheating behavior increased when participants were placed in a dark room (Zhong, Bohns, & Gino, 2010), were given limited time to think (Shalvi, Eldar, & Bereby-Meyer, 2012), when monetary rewards were given to charity (Lewis et al., 2012), or when others could benefit from their cheating behavior (Gino, Ayal, & Ariely, 2013).

While participants made the active decision to cheat or not, no significant differences emerged in Honesty-Humility between the self-initiated cheaters and fair players. The high cheating rate (75%) that was found in the current study, therefore, seems to be influenced by the situation, rather than a reflection of differences in personality traits. Since we created a fairly unambiguous and strong manipulation (i.e., an untraceable opportunity to cheat for a reward in an online setting), room for individual differences was limited. Hence, the obtained results are relatively uniform across participants. This concept of a *strong situation* (Ickes, 1982; Lissek, Pine, & Grillon, 2006; Mischel, 1977; Monson & Snyder, 1977), as described in social psychology, might be an important limitation to measuring interpersonal differences in the current experimental setting. A less anonymous situation in which participants are more hesitant to transgress experimental rules might reveal individual differences that did not emerge in the current study.

#### 4.3. *Applied implications*

The current study speaks to the ecological validity of the CIT. While the CIT has been tested in hundreds of studies, those studies typically used artificial settings, with trivial stimuli (e.g., play cards or mock crime items), undergraduate participants, little or no incentives, and an explicit instruction to cheat or lie. More recently, researchers have started to manipulate those factors exploring whether and how they influence CIT validity. Several factors have been found to have little or no affect, including stress and arousal during crime and test (Klein Selle et al., 2017; Peth, Vossel, & Gamer, 2012). Other factors, such as time between encoding and test (Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Nahari & Ben-Shakhar, 2011; Peth et al., 2012) have found to affect CIT validity. The current study adds to this growing body of research by indicating that self-initiated cheaters were indistinguishable in their responses from instructed cheaters.

Besides tackling the effect of self-initiated versus instructed cheating, the current paradigm also allows for testing incidentally acquired, real-world memories (see also Meixner & Rosenfeld, 2014) as opposed to controlled, artificial encoding of stimuli. That is, in most laboratory studies participants typically acquire information through mock crime procedures in which a certain object has to be stolen from a specified location. This results in an artificial focus on these stimuli, which have to be memorized for the sake of the experiment. In the current paradigm, participants acquired the critical information in a natural setting. By using various distracting filler tasks and a plausible cover story, it can be assumed that participants did not deliberately retain the information until the memory detection test. While this new method was particularly designed for external validity purposes rather than

demonstrating the CIT-effect, moderate to large effect sizes were found. These results are promising and should encourage further field application of memory detection.

#### 4.4. Limitations and suggestions for future research

This study is not without its limitations. First, the goal of the current paradigm was to compare self-initiated cheating with instructed cheating and fair playing, so it was crucial to elicit spontaneous cheating on the trivia quiz. It bears mentioning that some participants might have known the answer to the cheating-evoking questions and actually did not cheat. While the anonymous online setting cannot confirm cheating with absolute certainty, pilot studies in similar samples showed that the chance of knowing the answers to both questions was statistically very small. Still, in a subsequent experiment, this could be prevented by monitoring the participants' screen or asking non-existing questions of which the answers can only be found on a webpage designed by the experimenter.

Second, a question can be raised as to whether participants considered looking up the correct answers online as cheating, when the instructions merely required participants to solve the quiz "alone." Yet this wording was specifically chosen to create a situation in which participants could spontaneously initiate cheating behavior. Explicit instructions *not* to use Google were avoided, since it could have provoked the opposite effect, besides possibly revealing the true nature of the study. As a drawback of these instructions, participants might have assumed it was not necessarily prohibited. Interestingly, a substantial number of self-initiated cheaters (40%) felt the need to justify their correct response on the trivia quiz, when presented with an opportunity to confess. This might be an indication that using Google was in fact considered rule-breaking behavior.

Third, the study was run online. Participants used their own computer and Internet connection throughout the various stages of the 60-min experiment. While attentiveness during online testing has been voiced as a concern, various researchers have positively evaluated the use of online platforms for psychological studies (see Bartneck, Duenser, Moltchanova, & Zawieska, 2015; Mason & Suri, 2012). Despite fully acknowledging the lack of complete control, it is noteworthy that (1) attentiveness checks were inserted in the questionnaires and that very few participants failed the test; (2) incentives were included for good performance; (3) participants had extensive practice before commencing the CIT; and (4) strict exclusion criteria were used. Given the low number of participants who were excluded based on inattentiveness in the HEXACO questionnaire or in the CIT, as well as the high trial accuracy in the CIT for the remaining participants, we think that memory detection could be meaningfully investigated in the current online setting.

## 5. Conclusions

The results imply that the cognitive signature of lying (slower responses when lying as compared to truth telling) extends to self-initiated cheating. This finding is encouraging

from an ecological validity perspective and may pave the way for further field implementation of memory detection.

## Acknowledgments

This research was funded by a grant, no. 238/15, from the Israel Science Foundation to Gershon Ben-Shakhar and a grant from the University of Amsterdam to Bruno Verschuere. We thank Anouk Bercht for assistance in data collection.

## Notes

1. This criterion ensured that only those participants who understood the instructions and took the task seriously were included in the results (see Materials).
2. Pilot questionnaires revealed that the eight easy questions were each correctly answered by 88–100% of the sample, whereas the cheating-evoking questions were correctly answered by 0% (Who coined the term dinosaur?), 3% (Who wrote the autobiographical book *Wishful Drinking?*), and 6% (Who wrote the autobiographical book *Prairie Tale: A Memoir?*). Thus, the likelihood of answering two of these questions correctly is close to 0. When answering both cheating-evoking questions correctly, participants were labeled “cheaters.”
3. Participants were given an additional bonus when their individual CIT score ( $d_{\text{CIT}} = (M_{\text{RT}(\text{probes})} - M_{\text{RT}(\text{irrelevants})}) / SD_{\text{RT}(\text{irrelevants})}$ ) was below 0.2 (see also Noordraven & Verschuere, 2013).
4. The probe-irrelevant within-subject contrast was calculated as  $d_{\text{within}} = M(\text{RT}_{(\text{probes})} - \text{RT}_{(\text{irrelevants})}) / \sqrt{SD_{(\text{probes})}^2 + SD_{(\text{irrelevants})}^2 - 2 * r * SD_{(\text{probes})} * SD_{(\text{irrelevants})}}$ , where  $r$  is the Pearson correlation between  $RT_{(\text{probes})}$  and  $RT_{(\text{irrelevants})}$ . The between-subject contrast was calculated as  $d_{\text{between}} = (M_{\text{RT}(\text{probe-irrelevant difference group 1})} - M_{\text{RT}(\text{probe-irrelevant difference group 2})}) / \sqrt{((n_{(\text{group 1})} - 1) * SD_{(\text{probe-irrelevant difference group 1})}^2 + (n_{(\text{group 2})} - 1) * SD_{(\text{probe-irrelevant difference group 2})}^2) / (n_{\text{group1}} + n_{\text{group2}} - 2)}$ , see also Lakens (2013) and Suchotzki et al. (2017).

## References

- Abe, N., Suzuki, M., Mori, E., Itoh, M., & Fujii, T. (2007). Deceiving others: Distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. *Journal of Cognitive Neuroscience*, *19*(2), 287–295. <https://doi.org/10.1162/jocn.2007.19.2.287>.
- Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. <https://doi.org/10.1080/00223890902935878>.

- Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLoS ONE*, *10*(4), e0121595. <https://doi.org/10.1371/journal.pone.0121595>.
- Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Sciences*, *8*(5), 216–222. <https://doi.org/10.1016/j.tics.2004.03.012>.
- Blandón-Gitlin, I., Fenn, E., Masip, J., & Yoo, A. H. (2014). Cognitive-load approaches to detect deception: Searching for cognitive mechanisms. *Trends in Cognitive Sciences*, *18*(9), 441–444. <https://doi.org/10.1016/j.tics.2014.05.004>.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: The external validity of mock crime studies. *Journal of Experimental Psychology: Applied*, *9*, 261–269. <https://doi.org/10.1037/1076-898X.9.4.261>.
- Caso, L., Gnisci, A., Vrij, A., & Mann, S. (2005). Processes underlying deception: An empirical analysis of truth and lies when manipulating the stakes. *Journal of Investigative Psychology and Offender Profiling*, *2*(3), 195–202. <https://doi.org/10.1002/jip.32>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, MI: Lawrence Erlbaum. <https://doi.org/10.4324/9780203771587>.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. <https://doi.org/10.1007/BF02310555>.
- De Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO Personality Inventory: Psychometric properties, self-other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment*, *90*(2), 142–151. <https://doi.org/10.1080/00223890701845195>.
- DeAndrea, D. C., Carpenter, C., Shulman, H., & Levine, T. R. (2009). The relationship between cheating behavior and sensation-seeking. *Personality and Individual Differences*, *47*(8), 944–947. <https://doi.org/10.1016/j.paid.2009.07.021>.
- Debey, E., Ridderinkhof, R. K., De Houwer, J., De Schryver, M., & Verschuere, B. (2015). Suppressing the truth as a mechanism of deception: Delta plots reveal the role of response inhibition in lying. *Consciousness and Cognition*, *37*, 148–159. <https://doi.org/10.1016/j.concog.2015.09.005>.
- Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: An experimental investigation using ego depletion and goal neglect. *Acta Psychologica*, *140*(2), 133–141. <https://doi.org/10.1016/j.actpsy.2012.03.004>.
- Domnich, A., Panatto, D., Signori, A., Bragazzi, N. L., Cristina, M. L., Amicizia, D., & Gasparini, R. (2015). Uncontrolled web-based administration of surveys on factual health-related knowledge: A randomized study of untimed versus timed quizzing. *Journal of Medical Internet Research*, *17*(4), 94–104. <https://doi.org/10.2196/jmir.3734>.
- Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin and Review*, *17*(4), 486–491. <https://doi.org/10.3758/PBR.17.4.486>.
- Effron, D. A., Bryan, C. J., & Murnighan, J. K. (2015). Cheating at the end to avoid regret. *Journal of Personality and Social Psychology*, *109*(3), 395–414. <https://doi.org/10.1037/pspa0000026>.
- Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition*, *2*(1), 33–41. <https://doi.org/10.1016/j.jarmac.2013.02.002>.
- Furedy, J. J., Davis, C., & Gurevich, M. (1988). Differentiation of deception as a psychological process: A psychophysiological approach. *Psychophysiology*, *25*(6), 683–688. <https://doi.org/10.1111/j.1469-8986.1988.tb01908.x>.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex*, *13*(8), 830–836. <https://doi.org/10.1093/cercor/13.8.830>.
- Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization*, *93*, 285–292. <https://doi.org/10.1016/j.jebo.2013.04.005>.

- Gino, F., & Pierce, L. (2009). The abundance effect: Unethical behaviour in the presence of wealth. *Organizational Behaviour and Human Decision Processes*, 109(2), 142–155. <https://doi.org/10.1016/j.obhd.2009.03.003>.
- Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social and General Psychology Monographs*, 132(3), 197–214. <https://doi.org/10.3200/MONO.132.3.197-214>.
- Halevy, R., Shalvi, S., & Verschuere, B. (2014). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research*, 40(1), 54–72. <https://doi.org/10.1111/hcre.12019>.
- Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, 57, 72–88. <https://doi.org/10.1016/j.jrp.2015.04.003>.
- Ickes, W. (1982). A basic paradigm for the study of personality, roles, and social behavior. In W. Ickes & E. Knowles (Eds.), *Personality, roles, and social behavior* (pp. 305–341). New York: Springer Verlag.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed). Oxford, UK: Oxford University Press.
- Klein Selle, N., Verschuere, B. J., Kindt, M., Meijer, E. H., Nahari, T., & Ben-Shakhar, G. (2017). Memory detection: The effects of emotional stimuli. *Biological Psychology*, 129, 25–35. <https://doi.org/10.1016/j.biopsycho.2017.07.021>.
- Kleinberg, B., & Verschuere, B. (2015). Memory detection 2.0: The first web-based memory detection test. *PLoS ONE*, 10(4), e0118715. <https://doi.org/10.1371/journal.pone.0118715>.
- Kleinberg, B., & Verschuere, B. (2016). The role of motivation to avoid detection in reaction time-based concealed information detection. *Journal of Applied Research in Memory and Cognition*, 5(1), 43–51. <https://doi.org/10.1016/j.jarmac.2015.11.004>.
- Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, 58(8), 605–613. <https://doi.org/10.1016/j.biopsycho.2005.07.040>.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329–358. [https://doi.org/10.1207/s15327906mbr3902\\_8](https://doi.org/10.1207/s15327906mbr3902_8).
- Lee, K., & Ashton, M. C. (2012). Getting mad and getting even: Agreeableness and Honesty-Humility as predictors of revenge intentions. *Personality and Individual Differences*, 52(5), 596–600. <https://doi.org/10.1016/j.paid.2011.12.004>.
- Lewis, A., Bardis, A., Flint, C., Mason, C., Smith, N., Tickle, C., & Zinser, J. (2012). Drawing the line somewhere: An experimental study of moral compromise. *Journal of Economic Psychology*, 33(4), 718–725. <https://doi.org/10.1016/j.joep.2012.01.005>.
- Lieblisch, I., Kugelmass, S., & Ben-Shakhar, G. (1970). Efficiency of GSR detection of information as a function of stimulus set size. *Psychophysiology*, 6(5), 601–608. <https://doi.org/10.1111/j.1469-8986.1970.tb02249.x>.
- Lissek, S., Pine, D. S., & Grillon, C. (2006). The strong situation: A potential impediment to studying the psychobiology and pharmacology of anxiety disorders. *Biological Psychology*, 72(3), 265–270. <https://doi.org/10.1016/j.biopsycho.2005.11.004>.
- Luria, A. (1932). *The nature of human conflicts*. New York: Liveright. <https://doi.org/10.1097/00000441-193306000-00027>.
- 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>.
- Marston, W. M. (1920). Reaction-time symptoms of deception. *Journal of Experimental Psychology*, 3(1), 72–87. <https://doi.org/10.1037/h0067963>.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>.

- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644. <https://doi.org/10.1509/jmkr.45.6.633>.
- Meixner, J. B., & Rosenfeld, J. P. (2014). Detecting knowledge of incidentally acquired, real-world memories using a P300-based concealed-information test. *Psychological Science*, 25(11), 1994–2005. <https://doi.org/10.1177/0956797614547278>.
- Mischel, W. (1977). The interaction of person and situation. In D. Magnusson, & N. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333–352). Hillsdale, NJ: Erlbaum.
- Mohamed, F. B., Faro, S. H., Gordon, N. J., Platek, S. M., Ahmad, H., & Williams, J. M. (2006). Brain mapping of deception and truth telling about an ecologically valid situation: Functional MR imaging and polygraph investigation—initial experience. *Radiology*, 238(2), 679–688. <https://doi.org/10.1148/radiol.2382050237>.
- Monson, T. C., & Snyder, M. (1977). Actors, observers, and the attributional process: Toward a reconceptualization. *Journal of Experimental Social Psychology*, 13(1), 89–111. [https://doi.org/10.1016/0022-1031\(77\)90016-6](https://doi.org/10.1016/0022-1031(77)90016-6).
- Nagin, D. S., & Pogarsky, G. (2003). An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity. *Criminology*, 41(1), 167–194. <https://doi.org/10.1111/j.1745-9125.2003.tb00985.x>.
- Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, 48(6), 733–744. <https://doi.org/10.1111/j.1469-8986.2010.01148.x>.
- Nahari, T., Breska, A., Elber, L., Klein Selle, N., & Ben-Shakhar, G. (2017). The external validity of the Concealed Information Test: The effect of choosing to commit a mock crime. *Applied Cognitive Psychology*, 31(1), 81–90. <https://doi.org/10.1002/acp.3304>.
- Nathanson, C., Paulhus, D. L., & Williams, K. M. (2006). Predictors of a behavioral measure of scholastic cheating: Personality and competence but not demographics. *Contemporary Educational Psychology*, 31(1), 97–122. <https://doi.org/10.1016/j.cedpsych.2005.03.001>.
- Noordraven, E., & Verschuere, B. (2013). Predicting the sensitivity of the reaction time-based Concealed Information Test. *Applied Cognitive Psychology*, 27(3), 328–335. <https://doi.org/10.1002/acp.2910>.
- Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *NeuroImage*, 25(1), 267–277. <https://doi.org/10.1016/j.neuroimage.2004.10.041>.
- Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, 49(3), 381–390. <https://doi.org/10.1111/j.1469-8986.2011.01313.x>.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 16(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>.
- 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., 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>.
- Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, 23(10), 1264–1270. <https://doi.org/10.1177/0956797612443835>.
- Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: The scope and limits. *Trends in Cognitive Sciences*, 12(2), 48–53. <https://doi.org/10.1016/j.tics.2007.11.008>.
- Spence, S. A., Farrow, T. F., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport*, 12(13), 2849–2853. <https://doi.org/10.1097/00001756-200109170-00019>.



- 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>.
- Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.) (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge, UK: Cambridge University Press.
- 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. <https://doi.org/10.1002/acp.1601>.
- Verschuere, B., Crombez, G., Koster, E. H., Van Bockstaele, B., & De Clercq, A. (2007). Startling secrets: Startle eye blink modulation by concealed crime information. *Biological Psychology*, *76*(1–2), 52–60. <https://doi.org/10.1016/j.biopsycho.2007.06.001>.
- Verschuere, B., Spruyt, A., Meijer, E. H., & Otgaar, H. (2011). The ease of lying. *Consciousness and Cognition*, *20*(3), 908–911. <https://doi.org/10.1016/j.concog.2010.10.023>.
- Vrij, A. (2004). Guidelines to catch a liar. In P. Granhag & L. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 287–314). Cambridge, UK: University Press.
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253–265. <https://doi.org/10.1007/s10979-007-9103-y>.
- Vrij, A., Mann, S., Leal, S., & Fisher, R. P. (2010). ‘Look into my eyes’: Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime and Law*, *16*(4), 327–348. <https://doi.org/10.1080/10683160902740633>.
- Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research*, *22*(4), 544–562. <https://doi.org/10.1111/j.1468-2958.1996.tb00378.x>.
- Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774. <https://doi.org/10.1002/acp.914>.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>.
- Zhong, C. B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, *21*(3), 311–314. <https://doi.org/10.1177/0956797609360754>.