A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE aps

Research Report

Cheating the Lie Detector

Faking in the Autobiographical Implicit Association Test

Bruno Verschuere,[1] Valentina Prati,[2] and Jan De Houwer[1]

[1]Ghent University and [2]Padua University

ABSTRACT—*The autobiographical Implicit Association Test (aIAT) was recently introduced in this journal as a new and promising lie-detection tool. The initial report found 91% accuracy in determining which of two autobiographical events was true. It was suggested that the aIAT, unlike other lie-detection tests, is resistant to faking. We investigated whether participants can strategically alter their performance on the aIAT. Experiment 1 showed that participants guilty of a mock theft were able to obtain an innocent test outcome. Two additional experiments showed that guilty participants can fake the aIAT without prior experience with the aIAT and when a response deadline is imposed. The aIAT is subject to the same shortcomings as other lie-detection tests.*

In its war on terror, the U.S. government now uses handheld polygraphs for rapid screening of suspects of terrorism (e.g., "Are you a member of the Taliban?"). This illustrates the great need that exists among law-enforcement agencies for lie-detection tools that are easy to apply. The enthusiastic use of the pocket polygraph stands in sharp contrast to the highly critical evaluation of this method by the prominent National Research Council (2003), which concluded that errors frequently occur and that successful faking in polygraph tests is possible. An important recommendation was to develop new lie-detection methods.

The autobiographical Implicit Association Test (aIAT) is a very simple new lie-detection tool. The aIAT is based on reaction times and requires only a standard computer. The aIAT can be used to assess which of two autobiographical events is true. In an IAT used for criminal investigations, for example, sentences related to four categories are used: true statements unrelated to the crime (e.g., "I'm in front of a computer"), false statements unrelated to the crime (e.g., "I'm in the city library"), confession statements confirming that the participant committed the crime (e.g., "I stole the CD-ROM containing exam questions"), and denial statements denying that the participant committed the crime (e.g., "I did not steal the CD-ROM"). Sentences are presented one by one, and participants are required to give a speeded response depending on the task (see Table 1). The aIAT consists of two tasks. In the confession-true task, confession and true statements are mapped to one key, and denial and false statements are mapped to the second key. In the denial-true task, the assignments are reversed (denial and true statements are assigned to one key, confession and false statements assigned to the other key). Sartori, Agosta, Zogmaister, Ferrara, and Castiello, et al. (2008) argued that guilty participants should be faster in the confession-true task than in the denial-true task, whereas the reverse should be the case for innocent participants. In a series of six experiments, Sartori et al. found that the aIAT has an extremely high accuracy in determining which of two autobiographical events is true (overall 91%). The authors concluded that "the aIAT is an accurate method to detect concealed knowledge and outperforms currently available lie-detection techniques" (p. 780). This conclusion may be premature because it is unknown whether the aIAT is susceptible to faking.

In Experiment 1, we investigated whether participants can strategically alter their performance on the aIAT by providing them with a single instruction sheet on how to beat the aIAT. Subsequent experiments examined the conditions under which effects on the aIAT can be faked.

Address correspondence to Bruno Verschuere, Department of Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium, e-mail: bruno.verschuere@ugent.be.

EXPERIMENT 1

Method

Participants

Thirty-six undergraduate students at Ghent University were paid €4 for participation in this study. Eighteen participants (4 men, 14 women; mean age = 19.39 years, *SD* = 1.19 years) were assigned to the guilty condition, and 18 participants (4 men, 14 women; mean age = 19.55 years, *SD* = 1.20 years) were assigned to the innocent condition.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

TABLE 1
Response Assignment in the Autobiographical Implicit Association Test in This Experiment

	Required response	
Task	Press left key	Press right key
Confession-true	"I have stolen the CD-ROM" or true	"I did not steal the CD-ROM" or false
Denial-true	"I have stolen the CD-ROM" or false	"I did not steal the CD-ROM" or true

Note. Guilty participants were expected to be faster in the confession-true task than in the denial-true task. Innocent participants were expected to show the reverse pattern. In the confession-true task, participants were to press the left key if the statement was a confession or true statement, and press the right key if the statement was a denial or false statement. In the denial-true task, the assignments were reversed.

Stimuli and Procedure

All stimuli were similar to those used by Sartori et al. (2008, Experiment 2; see Table 2). They were presented on a 17-in. screen using Inquisit Software 3.0.1, which also recorded reaction times with millisecond accuracy.

The mock-crime procedure was a replication of the mock-crime procedure used by Sartori et al. (2008, Experiment 2). Participants chose one of two envelopes that assigned them to the guilty or innocent condition. All participants were instructed to leave the laboratory. Participants in the guilty condition went to a professor's office and stole a CD-ROM with the copy of an exam. Participants in the innocent condition read a newspaper article describing the theft. Upon their return to the laboratory, a first aIAT was administered. Participants pressed one of two keys to categorize target sentences as "true" or "false" and attribute sentences concerning the mock crime as indicating guilt ("I have stolen the CD-ROM") or innocence ("I did not steal the CD-ROM"; see Table 2). In confession-true task, "I have stolen the CD-ROM" and true sentences were assigned to the same response; in denial-true task, these sentences were assigned to different responses. Before executing the aIAT a second time, participants were given a sheet containing information on how the aIAT works and how to obtain an innocent test outcome. Reasoning that it would be easier for participants to slow down in the confession-true task than to speed up in the denial-true task, we explicitly instructed participants to slow down performance in the confession-true task (Fiedler & Bluemke, 2005).

Results and Discussion

The aIAT was scored using Greenwald, Nosek, and Banaji's (2003) D600 scoring algorithm. Positive scores indicate a stronger tendency to associate "I have stolen the CD-ROM" with true, hence a "guilty" test outcome. Negative scores indicate a greater tendency to associate "I have stolen the CD-ROM" with false, hence an "innocent" test outcome. As expected, guilty

TABLE 2
Sentences Used in the Autobiographical Implicit Association Test

Category	Sentence	Ground truth
True statement	I am in the Department of Psychology	True for all participants
	I am in a little room	
	I am taking part in an experiment	
	I am in the basement	
	I am in front of the computer	
False statement	I am in the city library	False for all participants
	I am in the bathroom	
	I am eating something in a road restaurant	
	I am at a tennis match	
	I am in a store	
"I have stolen the CD-ROM"	I went into the professor's office	True for guilty participants; false for innocent participants
	I stole the CD with the exam	
	I have stolen the clinical psychology exam	
	I went into the room to steal the CD	
	I certainly stole the exam	
"I did not steal the CD-ROM"	I never went into the professor's office	True for innocent participants; false for guilty participants
	I have never stolen the CD-ROM with the exam	
	I did not steal the clinical psychology exam	
	I did not go into the room for the CD	
	I certainly did not steal the exam	

Note. Sentences have been translated from the original Dutch.

TABLE 3

Mean Implicit Association Test Effects (D600s) and Hit Rates in the Three Experiments

Experiment and condition	Condition: Control D600	Control Hit rate (%)	Faking D600	Faking Hit rate (%)
Experiment 1				
Guilty	+0.30 (0.52)	67	−0.41 (0.61)	22
Innocent	−0.12 (0.52)	61	−0.34 (0.78)	72
Experiment 2				
Guilty, novice			+0.06 (0.61)	61
Guilty, experienced	+0.41 (0.37)	84	−0.43 (0.63)	26
Experiment 3				
Guilty, unspeeded	+0.29 (0.57)	76	−0.12 (0.52)	42
Guilty, speeded	+0.22 (0.25)	86	−0.09 (0.44)	38

Note. Positive D600 scores indicate a stronger tendency to associate ''I have stolen the CD-ROM'' sentences with true sentences, hence a ''guilty'' test outcome. Negative D600 scores indicate a greater tendency to associate ''I have stolen the CD-ROM'' sentences with false sentences, hence an ''innocent'' test outcome (Greenwald, Nosek, & Banaji, 2003). Standard deviations are given in parentheses.

participants obtained a more positive test score than innocent participants in the first aIAT, $t(34) = 2.57$, $p_{rep} = .96$, $d = 0.84$ (see Table 3). After receiving faking instructions, both guilty and innocent participants had negative test scores, $t(34) = .30$, $p_{rep} = .58$, $d = -0.10$.[1] In this second aIAT, the majority of guilty participants were erroneously classified as innocent. Guilty participants successfully altered their positive score into a negative score, $t(17) = 5.00$, $p_{rep} = .99$, $d = 1.17$, but faking instructions did not influence the scores of the innocent participants, $t(17) = 1.04$, $p_{rep} = .76$, $d = 0.25$.

An effective algorithm to detect fakers would render the problem of faking less serious. The challenge is not just to find an algorithm that detects fakers (high sensitivity), it is crucial that the algorithm does not falsely classify innocents (high specificity). No innocent control participant had a mean reaction time longer than 1,861 ms on the confession-true task, but six (33%) guilty fakers had longer mean reaction time. Thus, one could classify participants with a mean reaction time of more than 1,861 ms on the confession-true task as fakers. Because faking is not informative on guilt status, these participants can be excluded in accuracy calculation. After excluding identified fakers, specificity rose to 69%, but sensitivity remained low at 33%.

In sum, Experiment 1 showed that guilty participants were able to alter their aIAT test score to obtain a more innocent test outcome. Faking could not be detected in an obvious manner. Instructing innocent participants to obtain a more innocent test outcome, however, did not affect aIAT test scores.

[1]This effect was driven by pronounced slowing in the confession-true task. No reliable speeding was observed in the denial-true task. A similar dynamic was observed in Experiments 2 and 3. Because participants were explicitly instructed to slow down in the confession-true task, our data do not rule out the possibility that responding can be sped up in the denial-true task when instructions to do so are given.

EXPERIMENT 2

Fiedler and Bluemke (2005) have shown that participants may need prior experience with the IAT to be able to fake it. Chances are small that examinees in forensic contexts will have prior experience with the aIAT. However, everyone can easily gain experience with other versions of the IAT that are available on the internet. To examine the boundary conditions of the effects of faking in guilty participants, guilty participants in Experiment 2 had prior experience either with the aIAT (as in Experiment 1) or with the unrelated flower-insect IAT (Greenwald, McGhee, & Schwartz, 1998).

Method

Experiment 2 was identical to Experiment 1, except that there was no innocent condition, and half of the 36 guilty participants (2 men, 16 women; mean age = 19.39 years, $SD = 0.78$ years) first performed an unrelated flower-insect IAT (see Greenwald et al., 1998, Experiment 1), and half (1 man, 17 women; mean age = 20.28 years, $SD = 5.21$ years) performed the aIAT twice as in Experiment 1.

Results and Discussion

As in Experiment 1, guilty participants who had prior experience with the aIAT could strategically lower their test score after receiving faking instructions, $t(18) = 4.41$, $p_{rep} = .99$, $d = 1.01$. Table 3 shows that percentage accuracy dropped dramatically. The effect of faking was also found in guilty participants who first completed the flower-insect IAT. These participants had a lower score on the aIAT than the guilty participants when completing the aIAT for the first time, $t(35) = 2.11$, $p_{rep} = .92$, $d = 0.70$. The effect of faking was larger in participants who had prior experience with the aIAT compared to those who had prior experience with the flower-insect IAT, $t(35) = 2.41$, $p_{rep} = .95$, $d = 0.79$.

In sum, Experiment 2 again found faking in guilty participants who had prior experience with the aIAT, and showed that prior experience with the aIAT is not necessary for successful faking.

EXPERIMENT 3

Experiment 3 aimed at testing whether a response deadline could be used to prevent faking. Encouraging participants to speed up responding may make it more difficult to strategically slow down responding. Therefore, we introduced a response deadline that was set at about the mean reaction time in Experiment 1 (1,200 ms; also see Degner, in press).

Method

Experiment 3 was identical to the guilty condition in Experiment 1 for half of the 42 guilty participants (6 men, 15 women; mean age = 20.86 years, *SD* = 6.22 years). The other half (6 men, 15 women; mean age = 20.00 years, *SD* = 3.60 years) performed the aIAT under time pressure with the requirement to respond within 1,200 ms; for these participants, a "TOO SLOW" message was presented 1,200 ms after the start of the trial if no response was registered within that period.

Results and Discussion

Replicating results from Experiments 1 and 2, guilty participants could strategically lower their test score when there was no deadline, $t(20) = 3.51$, $p_{rep} = .99$, $d = 0.76$ (see Table 3). The effect of faking was also found in the speeded condition, $t(20) = 3.88$, $p_{rep} = .99$, $d = 0.84$. Percentage accuracy dropped in both conditions after participants received faking instructions. The aIAT scores in the speeded condition were not different from those obtained in the nonspeeded condition, $t < 1$, $p_{rep} < .65$. The response deadline appeared ineffective in countering faking in the aIAT.

GENERAL DISCUSSION

Our data support the validity of the aIAT in that guilty naive participants obtained a more positive ("guilty") test score compared to innocent participants. Hit rates in naive participants (67–86% in guilty and 61% in innocent participants) were, however, substantially lower than those obtained in the initial studies reported by Sartori et al. (2008). Given that laboratory research tends to overestimate accuracy because of reduced variability in characteristics related to test, participant, and context (National Research Council, 2003), these accuracy figures should be regarded as the upper boundaries of the aIAT's potential in forensic settings.

We also demonstrated that, like other lie-detection tests, the aIAT is not resistant to faking. After receiving faking instruction, a considerable percentage (39–78%) of the guilty participants was able to alter performance so as to obtain an innocent test outcome. Faking could not be detected in an obvious manner (Experiment 1) and could not be prevented using a response-deadline technique (Experiment 3).

Experiment 2 showed that prior experience with the aIAT helps to fake the test, but is not a necessary condition for successful faking. Participants who had completed a flower-insect IAT were also able to fake performance on the aIAT. Participants in our study may have been more likely to be familiar with the IAT than participants in forensic settings. Participants can, however, easily become familiar with the IAT through Web sites such as https://implicit.harvard.edu/implicit/demo/. Millions of people have already done so. Because these Web sites also give feedback, participants can train themselves up to the point that they can alter performance on the IAT. Familiarity and prior experience with the IAT are a major concern when implementing this test in forensic contexts.

There exists another reaction-time-based lie-detection test that is closely related to the IAT: the Timed Antagonistic Response Alethiometer (TARA; Gregg, 2007). The crucial difference with the aIAT is that TARA is based on a between-subjects comparison rather than a within-subjects comparison. There is only one critical block, which should be easier to complete for truth tellers than for liars. Successful faking in the TARA requires liars to speed up rather than slow down in the critical block. Because speeding up may be more difficult than slowing down, one could argue that TARA may be less vulnerable to faking than the aIAT. To test this hypothesis, new studies are needed in which liars are asked to fake by speeding up responding. Note that, because of its reliance on between-subjects comparisons, the TARA suffers from drawbacks that do not apply to the aIAT (see Sartori et al., 2008).

Acknowledgments—We thank Giuseppe Sartori and Sara Agosta for their assistance in setting up the autobiographical Implicit Association Test. Bruno Verschuere is a postdoctoral fellow of the Research Foundation–Flanders, Belgium (FWO).

REFERENCES

Committee to Review the Scientific Evidence on the Polygraph, National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: National Academies Press.

Degner, J. (in press). On the (un-)controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition & Emotion*.

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*, 307–316.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

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*, 197–216.

Gregg, A.P. (2007). When vying reveals lying: The timed antagonistic response alethiometer. *Applied Cognitive Psychology*, *21*, 621–647.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., & Castiello, U. (2008). How to accurately assess autobiographical events. *Psychological Science*, *19*, 772–780.

(RECEIVED 6/12/08; REVISION ACCEPTED 9/15/08)