In press in Psychological Science

How to accurately assess autobiographical events

Giuseppe Sartori[1], Sara Agosta[1], Cristina Zogmaister[2], Santo Davide Ferrara[3], Umberto Castiello[1,4]

[1]*Psychology Department, University of Padua, via Venezia 8, 35131 Padova, Italy*

[2]*Developmental and Social Psychology Department, University of Padua, via Venezia 8, 35131 Padova, Italy*

[3]*Environmental Medicine and Public Health Department, University of Padua, via Giustiniani 2, 35100 Padova, Italy*

[4]*Psychology Department, Royal Holloway-University of London, Egham Surrey, TW20EX*

Corresponding author:

Giuseppe Sartori, Via Venezia 8, 35100 Padova,

telephone number +390498276608

fax number: +390498276600

e-mail: giuseppe.sartori@unipd.it

Abstract

Here we describe a new method based on indirect measures of implicit autobiographical memory which allows to evaluate which of two contrasting autobiographical events (e.g., crimes) is true. Participants were requested to classify sentences describing autobiographical events by pressing one of two response keys. Results indicate that responses were faster when sentences related to real autobiographical events shared the same response key with other sentences reporting true events. In contrast, latencies were longer when real autobiographical events shared the same response key with sentences reporting false events. These results are discussed in the terms of a possible application of this method in forensic setting and as a lie detection technique.

Introduction

Autobiographical memory is the ability to remember events directly experienced by a person. The majority of studies on autobiographical memory rely on the extent of remembered information. (e.g., Crovitz & Schiffman, 1974). Furthermore, the evaluation of autobiographical memories, using indirect methods, has proved to be a useful indicator of guilty knowledge which may be valuable to detect lies (e.g., Lykken, 1960).

Detecting lies plays an essential role during crime investigations and criminal trials. Two of the most utilized tests for lie detection are the Control Question Test (CQT; Moore, Petrie & Braga, 2003) and the Guilty Knowledge Test (GKT; Lykken, 1960, 1998). The former is based on differential patterns of physiological activation (e.g., heart rate) accompanying direct questions addressed to the suspect (e.g., "*Did you do it?"*), as compared to neutral questions. The latter uses multiple-choice questions each including a "relevant" answer (e.g., feature of the crime under investigation) and several "control" answers which cannot be discriminated from the "relevant" answer by an innocent suspect (Lykken, 1998). Typically, in guilty suspects, larger physiological responses are detected for "relevant" rather than "control" alternatives. Recent developments in the use of these two tests consider their application within functional magnetic resonance imaging (fMRI)settings[1] (Ganis, Kosslyn, Stose, Thompson & Yurgelun-Todd, 2003; Langleben, Loughead, Bilker, Ruparel, Childress, Busch & Gur, 2005). However, despite such methodological advancements, these methods are still plagued by poor specificity and sensitivity (e.g., Iacono & Lykken, 1999).

[1] For a full account regarding the use of fMRI used for lie detection please refer to www.amacad.org/publications/bulletin/spring2007/science.pdf)

The use of the Implicit Association Test (IAT, Greenwald, McGhee & Schwarz, 1998) could provide an important step forward. For instance, Gray and colleagues elegantly illustrated how the IAT can be fruitfully applied in a forensic setting (Gray, Brown, MacCulloc, Smith & Snowden, 2005; Gray, MacCulloch, Smith, Morris & Snowden, 2003). They showed that it could correctly identify implicit beliefs in psychopathic murderers as well as pedophilic attitudes.

A further adaptation of the IAT which has the potential to be used in forensic setting is the Timed Antagonistic Response Alethiometer (TARA; Gregg, 2007). By means of response incongruity, TARA may be used to classify the respondent as a truth-teller or a liar on the basis of a speeded classification task of sentences.

Here we present a new IAT based methodology termed the Autobiographical IAT (aIAT). The aIAT allows one to evaluate which of two contrasting autobiographical events is true. This is accomplished by requiring participants to undergo two critical blocks of categorization trials, one of which should be facilitated if the respondent believes that one of the autobiographical events is true, whereas the other should be facilitated if the other event is believed to be true. The specific pattern of RT in the two blocks of trials will indicate the automatic assessment of truth for one of the events, and falseness for the other event.

General methods and procedures

The experimental procedures were approved by the Ethics Committee - University of Padua and were in accordance with the declaration of Helsinki.

Methods and procedure were similar for all experiments, except when specified. The computerized task consisted of five separate blocks of categorization trials. In each trial, a stimulus was presented at the center of the monitor. Participants were requested to classify it as fast and accurately as

possible, by pressing one of two labeled keys. Stimuli were sentences of variable length, each describing an autobiographical fact. In Block 1 (20 trials) participants classified sentences along the logical dimension "TRUE/FALSE". They pressed the "A" key if the sentence was of the "TRUE" type (e.g., "*I am in front of a computer"*) and the "L" key if the sentence was of the "FALSE" type (e.g., "*I am flying"*). In Block 2 (20 trials) participants classified sentences along the critical dimension "GUILTY/INNOCENT". They classified with the "A" button sentences of the "GUILTY" type (e.g., "*I made use of cocaine last month"*) and with the "L" button sentences of the "INNOCENT" type (e.g., "*I have never made use of cocaine"*). The specific categories for each of the experiments, which correspond to the general labels "GUILTY" and "INNOCENT", are reported in Table 2. In Block 3 (60 trials, double categorization block) they were requested to press the "A" key if the sentence was either of the "TRUE" type or of the "GUILTY" type, and the key "L", if the sentence was of the "FALSE" type or of the "INNOCENT" type. In Block 4 (40 trials) participants were requested to perform the inverse classification to Block 2. They pressed the "A" key for sentences for the "INNOCENT" type and the "L" key for sentences of the "GUILTY" type. In Block 5 (60 trials, double categorization block) participants pressed the "A" key for "TRUE" and "INNOCENT" sentences and the "L" key for "FALSE" and "GUILTY" sentences. Reminder labels in the form of category names remained on the monitor for the entire block duration. An error signal appeared when incorrect response occurred. Stimuli of the "TRUE" vs. "FALSE" type, and stimuli of the "GUILTY" vs. "INNOCENT" type were presented in alternate order in Blocks 3 and 5. Half of the participants were administered the blocks in this order, whereas for the other half the order of Blocks 3 and 5 was reversed (and the order of Blocks 2 and 4 was reversed accordingly). Preliminary analyses indicated that the order of presentation did not influence the main results and did not interact with the other factors. Therefore the order of presentation has been collapsed. Table 1 provides a schematic description of the aIAT.

The comparison of interest is between average RT in Block 3 and in Block 5. Both “Guilty” and “Innocent” respondents took part to Experiments 1, 2, and 5. The expected pattern of facilitation/inhibition should indicate that “Innocent” participants be faster in the block that associates “INNOCENT” sentences with “TRUE” (congruent block), as compared to the block that associates “GUILTY” sentences with “TRUE” (incongruent block), whereas the opposite should be revealed for “Guilty” participants. The specific pattern of facilitation should depend on each individual’s autobiographical knowledge. No “Innocent” participants were included in Experiments 3, 4, and 6: we expected all participants to be faster in the block of trials in which “GUILTY” sentences were associated with “TRUE”, as compared to the block in which “GUILTY” sentences were associated with “FALSE”.

Data analysis

Two dependent measures were considered: mean RT in the double-categorization blocks and D (Greenwald, Nosek, & Banaji, 2003). RT less than 150ms or longer than 10.000 ms were discarded. Unless specified, data have been submitted to an analysis of variance (ANOVA) with group (Guilty vs. Innocent) as a between subject factor and congruency (Congruent vs. Incongruent) as a within subject factor. The D index includes a penalty for incorrect trials, and expresses the IAT effect (the difference in performance between the two double-categorization blocks) in terms of the standard deviation of the latency measures. It is calculated by subtracting corrected mean RT for the block associating “INNOCENT” and “TRUE” sentences, from mean RT in the block associating “GUILTY” and “TRUE” sentences. Then, this difference was divided by the inclusive standard deviation of the two blocks. “Guilty” participants should show positive D values whereas “Innocent” participants should show negative D values. The number of participants correctly classified was computed by using the D score.

To discriminate between groups we conducted a Receiver Operating Characteristic (ROC) analysis (Swets, 1988). This analysis allows comparisons of the results obtained with the aIAT with those obtained with the GKT and fMRI methods (Ben-Shakhar & Elaad, 2003; Langleben et al., 2005).

Experiment 1. Cards

Here we evaluate the efficiency of the aIAT in identifying a selected card.

Participants. Thirty-seven students (19-30 yrs age; 8 males and 29 females) volunteered for the present study.

Procedure. Participants selected one of two playing cards ("*4 of diamonds*" and "*7 of clubs*") and memorized it in a preliminary consolidation task. In each trial of the consolidation task, participants were presented with one of eight different playing cards (e.g., 4 of diamonds, 7 of clubs, 3 of hearts, 3 of diamonds). They were asked to press the space bar if the card appearing centrally on the monitor was the chosen card. Each card was presented 5 times, for a total of 40 trials. An error feedback was presented for 400 ms if participants provided a wrong response. Out of the 37 participants 17 selected the "*4 of diamonds*" and 20 participants selected the "*7 of clubs*". After the consolidation task, participants performed the experimental task.

Stimuli. "GUILTY" sentences referred to the card selected by the participant while "INNOCENT" sentences referred to the non-selected card. Table 2 reports the list of "4 OF DIAMONDS" and "7 OF CLUBS" sentences. For Blocks 3 and 5 a total of 60 trials were presented (15 for "TRUE", 15 for "FALSE", 15 for "4 OF DIAMONDS" and 15 for "7 OF CLUBS" sentences). Each sentence was displayed until a response was emitted.

Results. Figure 1A represents the significant effect of congruency, $F(1,35)=37.275$, $p< .001$, $p_{rep} > .99$, $\eta^2=.516$. Mean latencies were lower for the congruent than for the incongruent block (972 vs. 1288 ms) suggesting a strong association between selected card and "TRUE" statements. No difference between choosers of "4 of diamonds" and "7 of clubs" emerged, $F(1,35)= 3.696$, $p= .063$, $\eta^2=.096$. The interaction between congruency and group was not significant, $F(1,35)= 0.459$, $p= .502$, $\eta^2=.013$.

To test the efficiency of the instrument, we computed the D index based on the difference between performance in the block associating the two categories "7 OF CLUBS" and "TRUE" and in the block associating "4 OF DIAMONDS" and "TRUE". Higher values of the index pointed to the autobiographical knowledge of having picked the "4 of diamonds", whereas lower values pointed to the opposite behavior. A positive mean D index emerged for the group who selected the "4 of diamonds", and a negative index for the group who selected the "7 of clubs" (.62 vs. -.49). This difference was significant, $F(1,35)= 82.753$, $p< .001$, $p_{rep}> .99$, $\eta^2= .70$. The accuracy of the method was confirmed by the ROC analysis (AUC = 0.985; see Figure 1B). The aIAT outperformed classification accuracy based on the GKT (Ben-Shakhar & Elaad, 2003; AUC = 0.80) and fMRI (Langleben et al., 2005; AUC = 0.80) for the same test. It accurately classified, using the D index, 35/37 participants.

Experiment 2. Mock Crime

Here "Guilty" participants simulated a theft, whereas "Innocent" participants simply read a press report on the same issue.

Participants. Thirty students volunteered for the experiment (14 males and 16 females; 23-30 yrs old; mean age=25.3). They were randomly assigned to the Guilty and Innocent groups.

Procedure and Stimuli. "Guilty" suspects were instructed to enter the office of a teaching assistant and steal a CD-ROM containing a to-be-done examination. "Innocent" suspects read a press report on this event. The aIAT procedure was similar to that used for Experiment 1, except for the "GUILTY" (e.g. *"I stole the CD-rom"*) and "INNOCENT" (e.g. *"I did not steal the CD-rom"*) sentences. The full list of "GUILTY" and "INNOCENT" sentences is reported in Table 2.

Results. Figure 1C shows that RT was faster for the congruent than for the incongruent condition (1091 vs. 1520 ms; $F(1,28)= 43.328$, $p< .001$, $p_{rep}> .99$, $\eta^2=.607$). RT for "Guilty" and "Innocent" participants did not differ, $F(1,28)=7.523$, $p= .011$, $\eta^2=.212$. The interaction between congruency and group was not significant, $F(1,28)= 3.892$, $p= .058$, $\eta^2=.122$.

Analysis of the D index revealed a significant difference between "Guilty" and "Innocent" participants (.78 vs. -.85; $F(1,28)=68.462$, $p< .001$, $p_{rep} > .99$; $\eta^2= .710$). All "Guilty" suspects showed a strong association between "GUILTY" sentences and "TRUE" sentences and therefore were correctly classified as guilty. Thirteen out of the 15 "Innocent" suspects showed a strong association between the "INNOCENT" and the "TRUE" sentences. The ROC analysis revealed an AUC= 0.96 (see Figure 1C). The aIAT outperformed the GKT (AUC=0.87; Ben-Shakhar & Elaad, 2003) in classification accuracy.

Experiment 3. Heroin and Cocaine

Here the aIAT is applied within an ecological setting: the detection of illegal behaviors such as drug usage.

Participants. Fourteen participants (13 males, 1 female; 23-45 yrs age; mean age=35.4) with at least 5 years of heroin and cocaine abuse were tested at a Local Substance Abuse Unit. Half of the participants were administered a version the aIAT that investigated their previous use of cocaine whereas the other half a version of the test that investigated their previous use of heroin.

Procedure and Stimuli. The "TRUE" and "FALSE" sentences were the same as for Experiments 1 and 2. The "GUILTY" sentences were concerned with heroin or cocaine usage, whereas the "INNOCENT" sentences were concerned with the non-usage of heroin and cocaine (see Table 2). The congruent condition consisted in "HEROIN(COCAINE)/TRUE" and "NON-HEROIN (NON-COCAINE)/FALSE" pairings whereas the incongruent condition consisted in "NON-HEROIN (NON-COCIANE)/TRUE" and "HEROIN(COCAINE)/FALSE" pairings.

Results. An ANOVA with group (Heroin vs Cocaine) as between-subject factor and congruency (Congruent vs Incongruent) as within-subject factor was conducted. No difference between participants responding to the Heroin-aIAT or to the Cocaine-aIAT emerged, $F(1,12)= .205$, $p=.659$, $\eta^2=.017$. The only significant effect indicated that responses to congruent associations were faster than responses to incongruent associations (1601 vs. 2234 ms; $F(1,12)=24.389$, $p< .001$, $p_{rep}> .99$, $\eta^2=.670$). No other effects approached significance. This pattern was also evident at individual level. The total number of drug users with a positive D was 13/14. The average D for the "Heroin" group was .98 and for the "Cocaine" group was .40.

Experiment 4. Autobiographical memory

It might be argued that the sentences used in Experiments 3 were not tapping into autobiographical memories, but rather describing participants' characteristics. To ascertain the efficiency of the aIAT

in detecting single autobiographical events limited in time and space, participants were asked to report a personal experience.

Participants. Twenty participants (8 males and 12 females, 19-53 yrs age) volunteered for the present study.

Procedure. To determine whether the aIAT could correctly identify the actual last vacation ("GUILTY" sentences) performed by the examinee, the critical association was between the actual last vacation with "TRUE" and a faked last holiday ("INNOCENT" sentences) with "FALSE". Participants were preliminarily requested to fill a questionnaire regarding their last vacations (e.g., *Last summer I went to New York*) and a vacation they never did (e.g., *Last summer I went to Kathmandu*). For each participant, a "personalized" aIAT was built with "GUILTY" sentences describing the true vacation and "INNOCENT" sentences describing a vacation which they never did (an example of the used sentences is reported in Table 2).

Results. Mean RT for the congruent block was faster than for the incongruent block (1041 vs. 1260 ms; $F(1,19)= 40.101$, $p< .001$, $p_{rep}> .99$; $\eta^2=0.679$). For 18/20 of the participants we correctly identified the real event based on the double-categorization block in which they were fastest. The average D score was .44.

Experiment 5. Suspension of driving license for drunk driving

A possible problem related to the previous experiments is that participants were not exposed to the high level of stress typical of an investigative setting and they would not experience direct advantages from faking. An important challenge for experimental studies of deception is to use a valid setting comparable to real situations where subjects may lie or conceal spontaneously.

Therefore we decided to run an experiment in which participants were highly motivated at passing the test. The main feature of the experimental group was that all participants had their driving license suspended for driving with excessive alcohol blood level.

Participants. Fifty participants (44 males and 6 females; 18-73 yrs age, mean age= 35.72) took part in the experiment. Twenty-five had their driving license suspended. A police control determined that all of them had, while driving, an alcohol blood level superior to 0.5 mg/ml. The remaining 25 participants were controls, matched to the experimental group for age, sex and education level and never caught while driving with an excess alcohol blood level (driving license track record).

Procedure and Stimuli. The aIAT was included as part of the compulsory medical and psychological assessment requested for the reinstatement of the driving license. Participants were made to believe that driving license reinstatement depended on the aIAT outcome. “TRUE” and “FALSE” sentences were the same as for all previous experiments. “GUILTY” sentences were 5 sentences describing the illegal act. “INNOCENT” sentences were sentences describing that the driver was never caught drunk by the police. The experimental group (“Guilty” participants) was expected to show an association between “TRUE” and “GUILTY” sentences (and between “FALSE” and “INNOCENT” sentences) whereas the control group (“Innocent” participants) was expected to show the reverse pattern. The full list of “GUILTY” and “INNOCENT” sentences is reported in Table 2.

Results. As shown in Figure 2A, RT for both groups (“Guilty” and “Innocent”) for the congruent block was faster than for the incongruent block (1805 vs. 2250 ms; $F(1,48)=32.029$, $p< .001$, $p_{rep}> .99$, $\eta^2= .400$). No other effect approached statistical significance. Analysis of the D index revealed that the difference between “Guilty” and “Innocent” participants was significant ($F(1,49)= 44.719$, $p< .001$, $p_{rep}> .99$ $\eta^2= .482$). The average D for the experimental group was positive whereas it was

negative for the control group (.39 vs. -.44). Using D, a total of 44/50 of the participants were correctly classified (22/25 for the experimental group and 22/25 for the control group). Finally the ROC analysis yielded an AUC= 0.91 (see Figure 2B).

Experiment 6: Criminals

We administered aIAT to two individuals who were found guilty after having confessed their crime and classified as mentally insane on the basis of a forensic psychiatric assessment. Both were under medication and were examined in a Forensic Mental Hospital. The first examinee (D.E.), attempted to kill his two sons. The second examinee (C.S.) was found guilty of killing his mother. For each criminal a personalized aIAT was built with "GUILTY" sentences describing the crime and "INNOCENT" sentences concerned with the denial of the crime.

Results. Administration of the aIAT to the first criminal (D.E.) revealed that he was faster for the congruent (4296 ms[2], 5 crime-related sentences such as "*I attempted to kill my children"* / "TRUE") than for the incongruent block (6733 ms, "*I did not attempt to kill my children"* / "TRUE"; $t(119) = -3.336$, $p < .001$, $p_{rep} > .99$; D= 1.0). This indicates a strong association between the "GUILTY" episode *"I attempted to kill my sons"* and the attribute "TRUE".

Administration of the aIAT to the second examinee (C.S.) revealed that average RT (1019 ms) for the congruent block (e.g., "GUILTY" sentence: "*I killed my mother"* / "TRUE") was significantly faster, $t(119) = –9{,}611$, $p < .001$, $p_{rep} > .99$, than for the incongruent block (2213 ms, e.g., "INNOCENT" sentences: "*I didn't kill my mother*" / "TRUE"). This pattern reveals a strong association between having *"killed my mother"* and "TRUE" sentences (D= .61).

[2] Very slow RT were presumably due to neuroleptic medication. Demonstration that the slowness in RT was not due to a specific 'faking' strategy is witnessed by a similar slowness detected when we administered another RT task (stop-signal) to this respondent.

Discussion

The present study reports a novel method allowing for a reliable detection of concealed autobiographical knowledge, which could be used in forensic science. Importantly, the Autobiographical IAT (aIAT) uses sentence-stimuli, rather than single words or pictures. It allows the investigation of autobiographical memory rather than semantic memory. The results from the experiments reported here provide compelling evidence of the high level of accuracy with which concealed autobiographical knowledge can be detected using this instrument. The aIAT provides a flexible and highly accurate method for detecting implicitly concealed knowledge. It is flexible because it can be used to submit the respondent with virtually any type of factual information in a verbal format. It is accurate because it can detect concealed knowledge not only at group level but also at individual participant level. On average, we were able to exactly classify 91% of the participants in a variety of differing tasks. Similarly to the GKT (Lykken, 1998) concealed knowledge measured with the aIAT could be used in lie detection.

A relevant issue is whether the aIAT can be faked. In this respect, Experiment 5 provides persuasive evidence that, even within an extremely faking-prone naturalistic setting, the aIAT is still able to accurately detect autobiographical event.

With respect to potential countermeasures, however, some studies report that IAT measures may be faked by participants who have been properly instructed to slow-down on compatible trials and to speed-up on incompatible trials (e.g., Fiedler & Bluemke, 2005; Kim, 2003; Steffens, 2004). The experiments reported here have been conducted with participants complying with the instructions. This condition is typical of “Innocent” suspects taking a “lie detection” test. There are also situations in which “Guilty” suspects accept to undergo the test that may prove their guilt. Naturally, when they do take the test, they are highly motivated to alter the results using appropriate

countermeasures such as those outlined by Fiedler & Bluemke (2005). Whether indirect indices (algorithms) could be developed to detect such countermeasures is an open issue for future research.

Issues concerned with faking are particularly evident with psychophysiological and neuroimaging techniques. In first instance, effective countermeasures to psychophysiological assessment are easy to implement. The polygraph may be faked if guilty suspects are trained in the use of physical (e.g. biting the tongue or pressing the toes to the floor) and mental countermeasures (e.g. engaging in mental activities that require effort such as counting backward; Honts, Raskin & Kircher, 1994; Ben-Shakar & Elaad, 2003). In second instance, although the use of fMRI-based techniques has revealed that activity within the frontal lobe is sensitive to the production and complexity of lies (e.g., Ganis et al., 2003), because of two main problems doubts were cast on the validity of this costly and cumbersome technique. First, the results could be faked by intentional head movements which may prevent an exact anatomical localization. Second, having guilty suspect to covertly engage in a concurrent cognitive task (such as backward counting) activates the "deception" frontal network (e.g., Cole & Schneider, 2007).

As a final issue, it might be said that both the aIAT and the TARA (Gregg, 2007) use response incongruity to identify lies. However, our method differs from the TARA in three important ways. First, TARA uses only two categories ("TRUE" and "FALSE") instead of the four ("TRUE" and "FALSE; "GUILTY" and "INNOCENT") used by aIAT. Second, in the application phase the TARA uses only one critical block instead of two congruent and incongruent blocks as for the typical IAT (Greenwald et al., 1998). Third, TARA discriminates truth from lie on the basis of an absolute level of RT on the critical block: if the average RT is fast then the respondent is honest, otherwise the respondent is lying. This procedure therefore requires a comparison with appropriate cut-offs obtained from carefully matched control groups. This may highlight a practical limit of the TARA. Consider the results obtained for the criminal (D.E.) tested in Experiment 6 of the present

paper. This criminal is a medicated patient with very slow RT. The use of the TARA in similar circumstances would require a medicated age-matched control group. Otherwise, using non-medicated controls with normal RT would cause a misclassification of the criminal as a liar even in the case he is responding truthfully.

To conclude, the aIAT is an accurate method to detect concealed knowledge which outperforms currently available lie-detection techniques. It can be used to assess the existence of virtually any kind of autobiographical memory in a range of malingered psychiatric and neurological disorders (e.g., malingered depression or malingered whiplash syndrome; Sartori, Agosta & Gnoato, 2007). All these aspects depict the potential of this method in providing novel insights for the detection of lies and malingering in forensic settings while opening important neuroethical issues (Wolpe, Foster & Langleben, 2005).

References

Ben-Shakhar, G. & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology*, *88*, 131-151.

Cole, M.W. & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, In Press.

Crovitz H.F. & Schiffman H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of Psychonomic Society*, 4, 517–18.

De Paulo, B.M. & Pfeifer, R.L. (1986). On the job experience and skill at detecting deception. *Journal of Applied Psychology*, *16*, 249-267.

Fiedler, K. & Bluemke, M., (2005). Faking the IAT: aided and unaided response control on the Implict Association Test. *Basic and Applied Social Psychology*, *27*, 307-316.

Ganis, G., Kosslyn, S.M., Stose, S., Thompson W.L. & Yurgelun-Todd, D.A. (2003). Neural correlates of different types of deception. *Cerebral Cortex*, *13*, 830-836.

Gray, N.S., MacCulloch, M.J., Smith, J., Morris, M. & Snowden, R.J. (2003). Violence viewed by psychopathic murderers. *Nature*, *423*, 497-498.

Gray, N.S., Brown, A.S., MacCulloch, M.J, Smith, J. & Snowden, R.J. (2005). An implicit test of the associations between children and sex in Pedophiles. *Journal of Abnormal Psychology*, 114,2, 304-308.

Gregg, A.I. (2007). When Vying Reveals Lying: The Timed Antagonistic Response Alethiometer. *Applied Cognitive Psychology*, *21*, 621-647.

Greenwald, A.G., McGhee, D.E. & Schwarz, J.L.K (1998). Measuring individual difference in implicit cognition: The implicit association test. *Journal of Personality & 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 & Social Psychology*, *85*, 197-216.

Honts, C.R., Raskin, D.C. & Kircher, J.C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, *79*, 252-259.

Iacono, W. G., & Lykken, D. T. (1999). Update: The scientific status of research on polygraph techniques: The case against polygraph tests. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony*. St. Paul, MN: West Publishing, pp. 174-184.

Kim, D.Y. (2003). Voluntary controllability of the Implicit Associatio Test. *Social Psychology Quartely*, *66*, 83-96.

Langleben, D.D., Loughead, J.W., Bilker, W.B., Ruparel, K., Childress, A.R., Busch, S.I. & Gur, R.C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *26*, 262-272.

Lykken, D.T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*, 258-262.

Lykken, D. (1998). *A tremor in the blood. The uses and abuses of the lie detector.* Reading, MA: Perseus Books

Moore H.M., Petrie, C.V. & Braga, A.A.(2003). *The polygraph and lie detection.* Washington DC. The National Academies Press.

Sartori, G., Agosta, S. & Gnoato, F., (2007). High accuracy detection of malingered whiplash sindrome. International Whiplash Trauma Congress. Miami, 26-27 October 2007.

Steffens, M.C. (2004). Is the Implicit Association Test Immune to Faking? *Journal of Experimental Psychology*, *51*, 165-179.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.

Wolpe, P.R, Foster, K.R. & Langleben, D.D. (2005). Emerging neurotechnologies for lie-detection: promises and perils. *The American Journal of Bioethics*, *5*, 39-49.

Figures captions

Figure 1. Results for Experiments 1 and 2. (**A**) Graphical representation for the interaction between group and stimulus pairings for Experiment 1. Participants who selected card "4 of diamonds" were faster when "GUILTY" sentences were paired with "TRUE" sentences and similarly participants who selected card "7 of clubs" were faster when "INNOCENT" sentences were paired with "TRUE" sentences. Congruent responses were faster than incongruent responses. (**B**) Representation of ROC for Experiment 1. The AUC approaching 1 indicates the high level of accuracy of the aIAT in identifying which card was selected by the participants. (**C**) Graphical representation for the interaction between group and stimulus pairings for Experiment 2. "Guilty" participants were faster when "GUILTY" sentences were paired with "TRUE" sentences, while "Innocent" participants were faster when "INNOCENT" sentences were paired with "TRUE" sentences. (**D**) Representation of ROC for Experiment 2. The AUC approaching 1 indicates the high level of accuracy of the aIAT in identifying "Guilty" and "Innocent" suspects. Sensitivity refers to the percentage of "Guilty" participants correctly classified as "Guilty". (1- Specificity) refers to the percentage of "Innocent" participants erroneously classified as "Guilty".

Figure 2. Results for Experiment 5. (**A**) Graphical representation for the interaction between groups and stimulus pairings for Experiment 5. Participants who had their driving license suspended because of drunken driving were faster when "GUILTY" sentences were paired with "TRUE" sentences while controls were faster when "INNOCENT" sentences were paired with "TRUE" sentences. (**B**) Representation of ROC for Experiment 5. The high AUC indicates the high level of accuracy of the aIAT in identifying drivers with suspended license and control participants. Sensitivity refers to the percentage of "Guilty" participants correctly classified as "Guilty". (1-Specificity) refers to the percentage of "Innocent" participants erroneously classified as "Guilty".

TABLE 1. Schematic description of the Autobiographical Implicit Association Test (aIAT) common to all experiments. The difference in average RT in Block 3 and Block 5 is used to identify an autobiographical event which is true for the respondent. If Block 3 is faster "GUILTY" sentences are true, if Block 5 is faster "INNOCENT" sentences are true for the respondent.

Sequence	Block 1	Block 2	Block 3	Block 4	Block 5
Task description	Logical categories discrimination	Initial autobiographical discrimination	Initial combined task	Reversed autobiographical discrimination	Reversed combined task
Response associated with "A" key	TRUE sentences "A" key	GUILTY sentences "A" key	TRUE/ GUILTY sentences "A" key	INNOCENT sentences "A" key	TRUE/ INNOCENT sentences "A" key
Response associated with "L" key	FALSE sentences "L" key	INNOCENT sentences "L" key	FALSE/ INNOCENT sentences "L" key	GUILTY sentences "L" key	FALSE/ GUILTY sentences "L" key

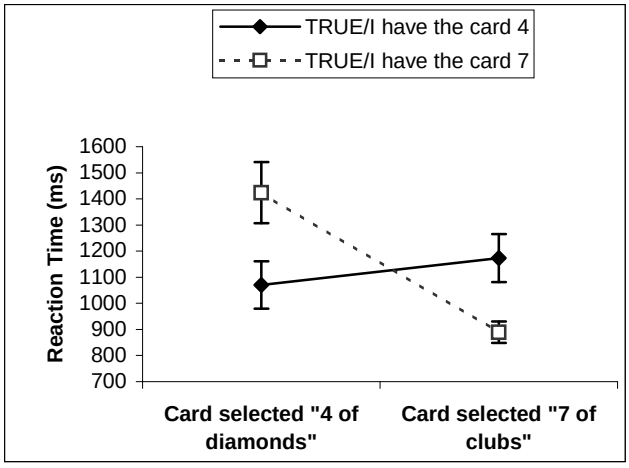
TABLE 2: List of sentences used for the five experiments. Please note that "TRUE" and "FALSE" sentences apply to all Experiments. The order for the remaining sentences follows the order of the Experiments (1-5).

Categories	English Translation	Comments
"True"	1. I'm in the basement of the Psychology department 2. I'm in a little room with a computer 3. I'm doing a psychology experiment 4. I'm in the laboratory of Psychology 5. I'm in front of the computer	Sentences certainly "True" for all the participants
"False"	1. I'm climbing a mountain 2. I'm at the beach 3. I'm eating in a downtown restaurant 4. I'm playing football 5. I'm in a shop	Sentences certainly "False" for all the participants
"4 of diamonds" (Exp. 1)	1. I picked card number 4 2. I turned card "four" 3. I saw the 4 of diamonds 4. I turned the 4 of diamonds 5. I have the 4 of diamonds	Sentences regarding the card "4 of diamonds"
"7 of clubs" (Exp. 1)	1. I picked card number 7 2. I turned card "seven" 3. I saw the 7 of clubs 4. I turned the 7 of clubs 5. I have the 7 of clubs	Sentences regarding the card "7 of clubs"
"I steal the CD-rom" (Exp. 2)	1. I entered in the professor's office 2. I stole a CD with the copy of the exam 3. I stole the exam of Clinical Neuropsychology 4. I entered in the office as to steal the cd-rom with the exam 5. I stole the exam	Sentences regarding the event "I steal the CD-rom"
"I did not steal the CD-rom" (Exp. 2)	1. I never entered in the professor's office to steal the cd-rom 2. I have never stolen the cd-rom containing the Clinical Neuropsychology exam 3. I did not steal the exam 4. I have never stolen the exam of Clinical Neuropsychology 5. I did not steal the exam of Clinical Neuropsychology	Sentences regarding the counter event "I did not steal the CD-rom"
"I used cocaine" (Exp. 3)	1. I have tried cocaine once 2. I took cocaine recently 3. I was addicted to cocaine 4. I used of cocaine 5. I was a cocaine abuser	Sentences regarding the event "I used cocaine "
"I did not use cocaine" (Exp. 3)	1. I never tried cocaine 2. I did not take cocaine 3. I was never addicted to cocaine 4. I never made use of cocaine 5. I was not a cocaine abuser	Sentences regarding the counter event "I did not use cocaine"

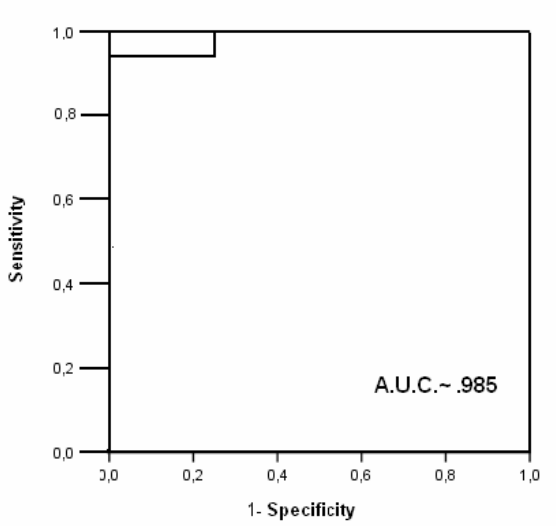
"I used heroine" (Exp. 3)	1. I have tried heroine once 2. I took heroine recently 3. I was addicted to heroine 4. I made use of heroine 5. I was an heroine abuser	Sentences regarding the event "I used heroine "
"I did not use heroine" (Exp. 3)	1. I never tried heroine 2. I did not take heroine 3. I was never addicted to heroine 4. I have never made use of heroine 5. I was not a heroine abuser	Sentences regarding the counter event "I did not use heroine"
"I went to Paris" (Exp. 4)	1. Last summer I went to Paris 2. I saw the Tour Eiffel 3. I visited the Louvre 4. I saw "The Monnalisa" 5. I visited the "Arc de Triomphe"	Sentences regarding the "Real" vacation
"I went to London" (Exp. 4)	1. Last summer I went to London 2. I saw the Big Ben 3. I had a typical English breakfast 4. I visited Tate modern Museum 5. I visited the British museum	Sentences regarding the "False" vacation
"My driving license was suspended because of alcohol" (Exp. 5)	1. I drove after I drank, thus my driving license was suspended 2. I drove my car while drunk, and they suspended my driving license 3. I drove while not sober, and they suspended my driving license 4. They suspended my driving license because I was drunk and I was driving 5. They suspended my driving license because I was above the alcohol level.	Sentences regarding the event "My driving license was suspended because of alcohol "
"My driving license was not suspended because of alcohol" (Exp. 5)	1. My driving license was not suspended because I was drunk 2. They did not suspended my driving license because of alcohol level 3. My driving license was not suspended because I was above the alcohol level 4. They never suspended my driving license because I was drunk 5. They never suspended my driving license because I was above the threshold of alcohol	Sentences regarding the counter event "My driving license was not suspended because of alcohol level"

Fig 1.

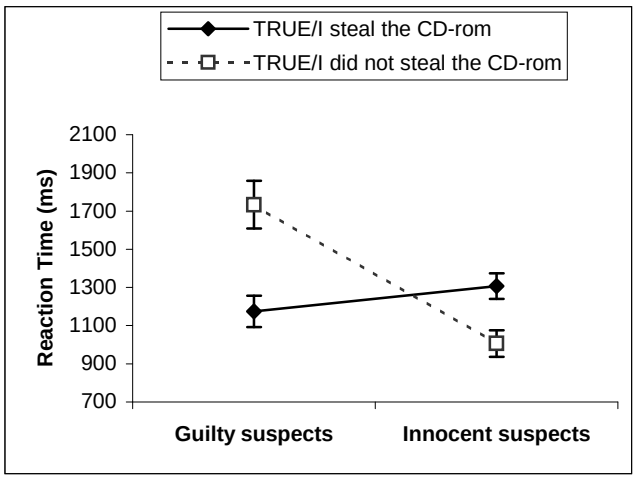
A


TRUE/I have the card 4
TRUE/I have the card 7
Reaction Time (ms)
1600
1500
1400
1300
1200
1100
1000
900
800
700
Card selected "4 of diamonds"
Card selected "7 of clubs"

B


Sensitivity
1,0
0,8
0,6
0,4
0,2
0,0
A.U.C.~ .985
0,0
0,2
0,4
0,6
0,8
1,0
1- Specificity

C


TRUE/I steal the CD-rom
TRUE/I did not steal the CD-rom
Reaction Time (ms)
2100
1900
1700
1500
1300
1100
900
700
Guilty suspects
Innocent suspects

D

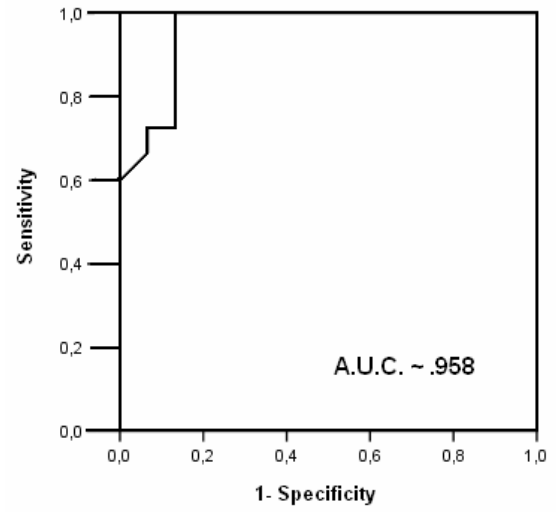
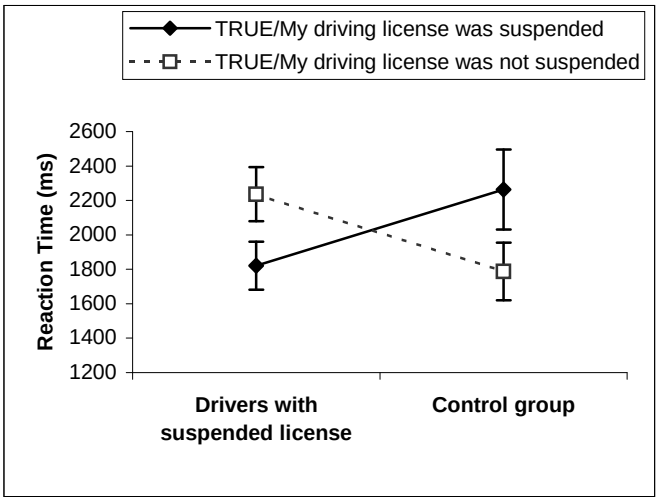
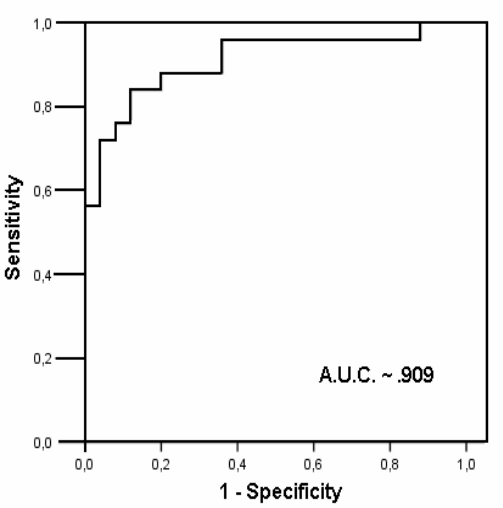

Sensitivity
1,0
0,8
0,6
0,4
0,2
0,0
A.U.C. ~ .958
0,0
0,2
0,4
0,6
0,8
1,0
1- Specificity

Fig 2

A


TRUE/My driving license was suspended
TRUE/My driving license was not suspended
Reaction Time (ms)
2600
2400
2200
2000
1800
1600
1400
1200
Drivers with suspended license
Control group

B


Sensitivity
1,0
0,8
0,6
0,4
0,2
0,0
A.U.C. ~ .909
0,0
0,2
0,4
0,6
0,8
1,0
1 - Specificity