Università degli Studi di Padova

Dipartimento di Psicologia dello Sviluppo e della Socializzazione

Facoltà di Psicologia

Scuola di Dottorato di Ricerca in Psicologia

Indirizzo in Scienze Cognitive

XXII ciclo

THE AUTOBIOGRAPHICAL IAT: A NEW TECHNIQUE FOR MEMORY DETECTION

Direttore della Scuola : Ch.ma Prof.ssa Clara Casco

Coordinatore d'indirizzo: Ch.ma Prof.ssa Anne Maass

Supervisore : Ch.mo Prof. Giuseppe Sartori

Dottorando : Sara Agosta

31 Dicembre 2009

INDEX

ABSTRACT

The autobiographical IAT (aIAT) is a new technique of memory detection that can be used to identify which of two autobiographical events is true. The technique is based on a classification task. Participants have to classify items in four different categories using only two motor responses. The underlying assumption is that the condition in which two associated concepts require the same motor response (congruent block) reaction times will be faster than the condition where the two associated concepts require two different motor responses (incongruent block).

The practical applications of this technique to the forensic field are straightforward. Six validation studies have been run (Chapter 2). In all the six experiments the true autobiographical event has been identified on the basis of the pattern of reaction times (RTs) in fact the congruent block show faster RTs than the incongruent block. It has also been shown that coached participants can successfully fake the aIAT, but faking can be detected on the basis of a specific pattern of reaction times (Chapter 3).

The accuracy and validity of the aIAT has been evaluated further and I showed that to enhance the validity of the instrument is necessary to be cautious in using sentences to describe autobiographical events (Chapter 4).

Finally, it has been shown that the aIAT can be applied to the identification of intentions, other than autobiographical events (Chapter 5). The application of the aIAT to the intention detection has been investigated also with Event Related Potentials (ERPs). Results showed a reduced Late Positive Component (LPC) for the incongruent block in respect to the congruent one. The LPC has been shown to be related to the cognitive control indicating here a stronger cognitive control during the incongruent than the congruent block.

In sum, the aIAT has been shown to be a reliable method that can be used to identify an autobiographical event or a future intention.

BREVE RIASSUNTO DEL LAVORO SVOLTO

Obiettivo della ricerca è stato quello di validare un nuovo strumento, l'Autobiographical Implicit Association Test (aIAT), basato su misure indirette, il cui scopo è verficare la veridicità di un evento autobiografico.

L'aIAT, stabilisce l'associazione fra la descrizione verbale di un evento (e un contro-evento) e la dimensione logica vero-falso. Compito del soggetto è quello di classificare delle frasi, che vengono presentate al centro di uno schermo, il più velocemente possibile. Ai partecipanti vengono presentati in ordine casuale item relativi a quattro concetti, due concetti target (evento-contro evento) e la dimensione logica (vero-falso); il compito dei partecipanti è quello di classificare gli item mediante due tasti; le risposte possibili del soggetto sono solamente due, in modo tale che i quattro concetti vengano associati a coppie.

L'assunto di base prevede che i partecipanti siano più veloci nel compito di classificazione quando i concetti associati richiedono la medesima risposta (compito congruente). Al contrario, quando i due concetti associati richiedono risposte differenti, i tempi di reazione saranno relativamente più lenti (compito incongruente). Tale procedura è stata validata mediante una serie di esperimenti (Capitolo 2) il cui scopo era discriminare:

-quale fra due carte è stata scelta da un partecipante, aIAT carte.

-fra due gruppi di partecipanti, coloro che hanno commesso un crimine da partecipanti che hanno letto un articolo di giornale, aIAT mock crime

-fra soggetti a cui è stata ritirata la patente per guida in stato d'ebbrezza e soggetti a cui non è mai stata ritirata la patente, aIAT guida in stato d'ebbrezza.

I tempi di reazione delle situazioni congruenti sono significativamente più veloci dei tempi di reazione delle situazioni incongruenti.

Studi successivi (Capitolo 3) hanno dimostrato come sia possibile utilizzare delle contromisure al test, ma queste stesse contromisure possono essere facilmente rintracciabili sulla base del pattern di tempi di reazione.

Un ulteriore milgioramento del test è stato effettuaato attraverso alcuni studi (Capitolo 4) che hanno dimostrato che l'utilizzo di frasi ed etichette nella forma negativa riduca l'accuratezza del test.

Infine l'aIAT è stato applicato allo studio delle intenzioni (Capitolo 5) e si è dimostrato in grado di individuare non solo gli eventi autobiografici accaduti in passato, ma anche le intenzioni future. Lo studio delle intenzioni è stato effettuato anche mediante la tecnica dei potenziali evocati che ha identificato una diversa componente tardiva (late positive component, LPC) nel blocco congruente ed incongruente, in particolare si è trovata una LPC ridotta nel blocco incongruente rispetto a quello congruente. La riduzione della LPC è stato associato in letteratura al controllo cognitivo, indicando quindi la necessità di un maggiore controllo mentre il partecipante svolge il blocco incongruente.

CHAPTER 1

"*In time of war the truth is so precious*

it must be attended by a bodyguard of lies" (Churchill)

DETECTING DECEPTION & AUTOBIOGRAPHICAL MEMORIES

One of the most common human behaviors is deception. Despite the wide use of deception in everyday life there seems not to be a uniquely recognized and accepted definition of deception. This problem is increased by the necessity of defining different types of lies (from an easy yes/no answer to a complicated constructed lie; e.g. Ganis, Kosslyn, Stose, Thompson & Yurgelun-Todd, 2003).

St. Augustine defined deception as an "intentional negation of a subjective truth" (Augustine, 1949). Mitchell defined deception as "any phenomenon which fulfill these three criteria: i) an organism R register (or believes) something Y from some organism S, where S can be described as benefiting when (or desiring that) iia) R acts appropriately toward Y, because iib) Y means X; and iii) and it is untrue that X is the case" (Mitchell, 1986). Mitchell distinguishes between 4 levels of deception (Mitchell, 1986):

1) At the first level the organism "acts" deceiving because he cannot do otherwise;
2) At the second level the organism deceive "doing p given that q";
3) At the third level deception is the result of an open program and this program can be modified by the result of the action of the organism;
4) At the fourth level there is an open program that is able to programming and reprogramming itself based upon the past and present actions.

Johnson and colleagues (Johnson, Barnhardt & Zhu, 2003) highlighted the common substrate to all definitions: "regardless of the nature and extent of the cognitive/emotional processes that precede and accompany a decision to deceive, all deceptions require the execution of a response that is incompatible with the truth". This definition highlights the core problem related to deception: the "incompatibility with the truth", on this base I developed a new lie detector based on implicit associations. Before entering in the specificity of the new method here it’s central to highlight the importance of deception from the phylogenetic and ontogenetic point of view and briefly describe previous methods that have been used to detect liars.

From Byrne’s (1996) point of view: “intelligence is an adaptation to deal with the complexity of living in semi-permanent groups of conspecifics, a situation that involves a complex tricky balance of competition and cooperation”. Social interactions involve an element of deception, thus the possibility of deceptive, well calculated communications and the necessity of detecting such machinations and manipulations provided a major impetus for the evolution of primate and human intelligence (Byrne, 1996). Byrne and Corp (2004) also showed that the use of deception within the primates is well predicted by the neocortical volume. This social function of intellect may be considered a primary context for his ontogenesis as well (LaFreniere, 1988).

Given the wide use of deception in everyday life it is important to study the markers of deception, in order to identify lies. Here I will present a brief description of the techniques used to spot lies and liars. It is possible to distinguish between: i) behavioral markers of deception (posture, response patterns, time reactions); ii) physiological markers of deception (heart beat, skin conductance, blink) and iii) brain related markers of deception (Dorso-Lateral Prefrontal Cortex, Anterior Cingulate).

Regarding the behavioral markers of deception, studies concentrated on the ability of identifying behaviors associated to deception. Non-verbal markers comprehend expressions of the face, bearings, gestures, eye blinks, and eyes movements. No clear relation has been shown between deception and these behavioral markers. De Paulo and colleagues (DePaulo, Lindsay, Malone, Muhlenbruch, Charlton & Cooper; 2003) showed, in their meta-analysis, a low connection between non-verbal markers and detection of deception.

Other non-verbal markers of deception concentrated on the study of the language style and paralinguistic expressions. Language styles refer to speech construction: use of ambiguous words (e.g., maybe, perhaps) or un-personalization word (limited use of personal hints). De Paulo and colleagues (2003) showed that when people are lying they are evasive, not clear and un-personal.

Physiological markers of deception are universally recognized as the traditional “lie detectors”. The most famous and used is best known as the *polygraph* (Ben-Shakhar & Elaad, 2003). The actual accuracy of the polygraph is a matter of controversy. Some researchers believe that the current system is not better than chance (e.g., Ben-Shakar, 1991). Other researchers estimate accuracy at 75% to 80% (e.g., Patrick & Iacono, 1991). This difference can be explained by the use of different procedures. 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, 1959, 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 electroencephalogram recording (EEG) (e.g., Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987; Farwell & Donchin, 1991) or functional magnetic resonance imaging (fMRI) settings (e.g., Ganis, et al., 2003; Langleben, Loughead, Bilker, Ruparel, Childress, Busch & Gur, 2005).

These last studies using fMRI identify the brain related markers of deception. In particular the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulated cortex (ACC) are activated during deception. The DLPFC is activated during working memory tasks and is involved in the necessary control of the outcome of the lie. The ACC is involved in the inhibition of the true response in favor of a false answer. However, despite such methodological advancements, these methods are still plagued by poor specificity and sensitivity determined by the use of CQT and GKT(e.g., Iacono & Lykken, 1999). Firthermore DLPFC and ACC cannot be considered as specific markers of deception, in fact they are activated also during other tasks (e.g. decision making involving inhibition and control, working memory)

Here I present a new technique that allows identifying which of two autobiographical events is true. It can be used for any autobiographical memory and has a great potential in forensic applications.

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). Here an application of the Implicit Association Test (IAT; Greenwald, McGhee & Schwarz, 1998) is used to identify which of two autobiographical events is true.

In Chapter 2 I will present a series of validation studies of the technique. In Chapter 3 I will deal with the important issue related to the faking of the test. In Chapter 4 I will present new studies aimed at enhancing the validity and accuracy of the new technique presented here and finally in Chapter 5 I will describe a further application of the test, in fact it will be applied not only to autobiographical information but also to intentions.

CHAPTER 2

VALIDATION STUDIES OF THE *AUTOBIOGRAPHICAL* IAT

INTRODUCTION

The Implicit Association Test (IAT; Greenwald et al., 1998) is at present one of the most used instruments, in psychology, to measure automatic implicit associations. In accord to Greenwald (Greenwald et al., 1998) the IAT measures the association strength between two concepts and a bipolar attribute. In the IAT, items related to four concepts are presented in a random order, participants have to classify these items in the respective category but, instead of requiring four kinds of responses, only two motor responses are needed (i.e. pressing two keys of the keyboard). The underlying assumption rely on the fact that if two strictly associated concepts require the same response individuals will be faster than if the two closely associated concepts require different responses. The IAT effect is defined as the difference in reaction times in two tasks: the task that requires two different responses for the two associated concepts and the task where the associated concepts require the same response. The IAT effect is considered an indicator of the degree of association between concepts.

The use of the Implicit Association Test (IAT, Greenwald et al., 1998) could provide an important step forward for identifying lies when used in the forensic setting. For instance, Gray and colleagues elegantly illustrated how the IAT can be fruitfully applied in a forensic setting to psychopaths and pedophiles (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 it is presented a new IAT based methodology termed 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 RTs 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 for all the experiments presented here 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 separated blocks of categorization trials. In each trial, a stimulus was presented in 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 event. 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 climbing a mountain"*). 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 2.2. 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 2.1 provides a schematic description of the aIAT.

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	GUILTY sentences	TRUE/ GUILTY sentences	INNOCENT sentences	TRUE/ INNOCENT sentences
Response associated with "L" key	FALSE sentences	INNOCENT sentences	FALSE/ INNOCENT sentences	GUILTY sentences	FALSE/ GUILTY sentences

Table 2.1: Schematic description of the autobiographical Implicit Association Test (aIAT) common to all experiments. The difference in average RTs 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.

The comparison of interest is between average RTs in Block 3 and in Block 5. Both guilty and innocent respondents took part to Experiments 2.1, 2.2, and 2.5. The expected pattern of facilitation/inhibition should indicate innocent participants to be faster in the block pairing Innocent sentences with True sentences (congruent block), as compared to the block that associates Guilty sentences with True sentences (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 2.3, 2.4, and 2.6: I expected all participants to be faster in the block of trials in which Guilty sentences were associated with True sentences, as compared to the block in which Guilty sentences were associated with False sentences.

GENERAL METHODS FOR DATA ANALYSIS

Two dependent measures were considered: average RTs in the double-categorization blocks and the D-IAT (Greenwald, Nosek, & Banaji, 2003). RTs 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-IAT includes a penalty for incorrect trials, and expresses the aIAT effect (the difference in performance between the two double-categorization blocks) in terms of the standard deviation of the latency measures. Here it is calculated by subtracting corrected mean RTs for the block associating Innocent and True sentences, from mean RTs 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 a Receiver Operating Characteristic (ROC) analysis that is expressed in terms of Area Under the Curve (AUC; Swets, 1988) have been conducted. The AUC is a measure of discrimination which ranges between 1, perfect discrimination, and 0.5, random discrimination. 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 2.1. Two-cards aIAT

Here the efficiency of the aIAT in identifying a selected card has been evaluated.

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 or7 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 card while 20 participants selected the 7 of clubs card. 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 2.2 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 2.1A represents the significant effect of congruency, (*F*(1,35)=37.275, *p*< .001, η^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 4 of diamonds or 7 of clubs choosers emerged, (*F*(1,35)= 3.696, *p*=.063, η^2=.096). The interaction between congruency and group was not significant, (*F*(1,35)= 0.459, *p*=.502, η^2=.013).

To test the efficiency of the instrument, I computed the D-IAT 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 sentences. 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-IAT 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, η^2=.70). The accuracy of the method was confirmed by the ROC analysis (AUC = 0.985; see Figure 2.1B). The aIAT outperformed classification accuracy based on the GKT (Ben-Shakhar & Elaad, 2003; AUC = 0.80) and fMRI (Langleben et al., 2005; AUC =.80) for the same test. It accurately classified, using the D index, 35/37 participants.

Experiment 2.2. Mock-Crime aIAT

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 2.1, 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 2.2.

Results

Figure 2.1C shows that RTs were faster for the congruent than for the incongruent condition (1091 vs. 1520 ms; $F(1,28)= 43.328$, $p<.001$, $\eta^2=.607$). RTs 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; $F(1,28)=68.462$, $p<.001$, $\eta^2=.710$). 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, therefore they were correctly classified as “Innocent”. The

ROC analysis revealed an AUC= .96 (see Figure 2.1D). The aIAT outperformed the GKT (AUC=.87; Ben-Shakhar & Elaad, 2003) in classification accuracy.

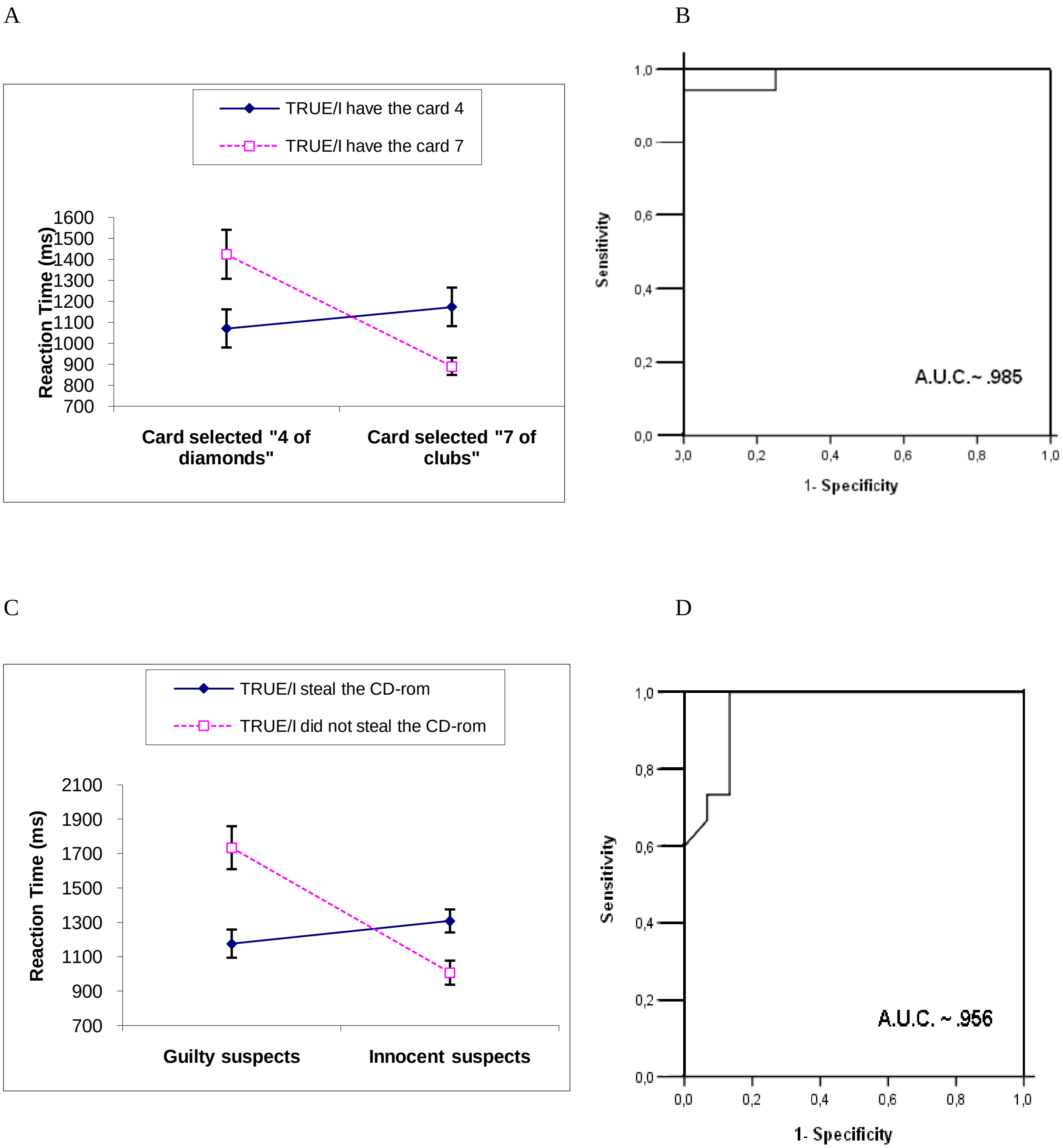


Figure 2.1: Results for Experiments 2.1 and 2.2. (**A**) Graphical representation for the interaction between group and stimulus pairings for Experiment 2.1. Participants who selected the 4 of diamonds card were faster when Guilty sentences were paired with True sentences and similarly participants who 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 2.1. 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 2.2. Guilty participants, that enacted the mock-crime, 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 2.2. 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”.

Experiment 2.3. Heroin and Cocaine aIAT

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 2.1 and 2.2. 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 2.2). The congruent condition consisted in “Heroin(Cocaine)/True” and “Non-Heroin (Non-Cocaine)/False” pairings whereas the incongruent condition consisted in “Non-Heroin (Non-Cocaine)/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 ms vs. 2234 ms; $F(1,12)=24.389$, $p<.001$, $\eta^2=.670$). No other effects approached significance. This pattern was also evident at individual level. The total number of drug users with a positive D-IAT was 13/14. The average D for the heroin group was .98 and for the cocaine group was .40.

Experiment 2.4. Autobiographical memory

It might be argued that the sentences used in Experiments 2.3 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 Los Angeles*). 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 2.2).

Results

Mean RTs for the congruent block was faster than for the incongruent block (1041 ms vs. 1260 ms; $F(1,19)= 40.101$, $p<.001$, $\eta^2=.679$). For 18/20 of the participants' the real event has been correctly identified, on the basis of the double-categorization block in which they were fastest. The average D score was .44.

Experiment 2.5. 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 I 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 educational 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 2.2.

Results

As shown in Figure 2.2A, RTs for both groups (guilty and innocent) for the congruent block were faster than for the incongruent block (1805 ms vs. 2250 ms; $F(1,48)=32.029$, $p<.001$, $\eta^2=.400$). No other effect approached statistical significance. Analysis of the D-IAT revealed that the difference between guilty and innocent participants was significant (*F*(1,49)= 44.719, *p*<.001, η^2=.482). The average D for the experimental group was positive whereas it was negative for the control group (.39 vs. -.44). Using the D-IAT, 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=.91 (see Figure 2.2B).

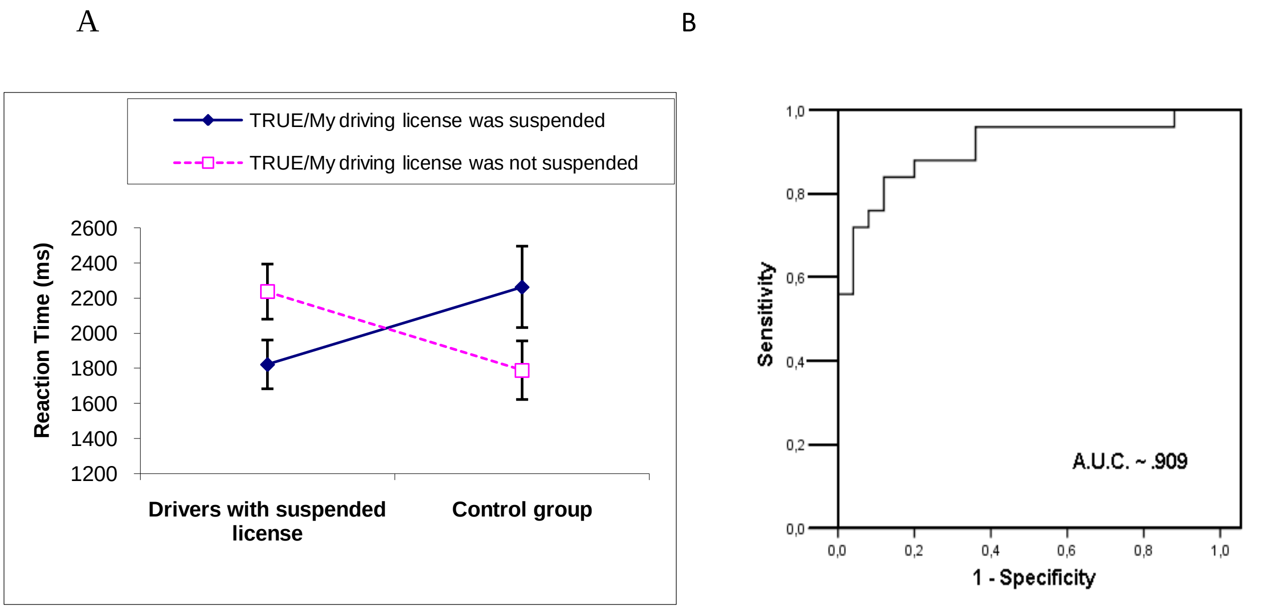


Figure 2.2: Results for Experiment 2.5. (A) Graphical representation for the interaction between groups and stimulus pairings for Experiment 2.5. Participants who had their driving license suspended because of drunken driving were faster when Guilty sentences were paired with True sentences while controls were faster when Innocent sentences were paired with True sentences. (B) Representation of ROC for Experiment 2.5. 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".

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	Sentences certainly "False" for all

	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	the participants
“4 of diamonds” (Exp. 2.1)	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. 2.1)	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. 2.2)	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. 2.2)	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. 2.3)	1. I have tried cocaine once 2. I took cocaine recently 3. I was addicted to cocaine 4. I used of cocaine	Sentences regarding the event “I used cocaine ”

	5. I was a cocaine abuser	
“I did not use cocaine” (Exp. 2.3)	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. 2.3)	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. 2.3)	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. 2.4)	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. 2.4)	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”	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	Sentences regarding the event “My driving license was suspended because of alcohol ”

(Exp. 2.5)	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.	
“My driving license was not suspended because of alcohol” (Exp. 2.5)	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”

Table 2.2: 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 (2.1-2.5).

Experiment 2.6: Criminals

I 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[1], 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) = -

[1] Very slow RTs were presumably due to neuroleptic medication. Demonstration that the slowness in RTs 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.

3.336, *p*<.001; D=1.0). This indicates a strong association between the Guilty sentences *I attempted to kill my sons* and the attribute True.

Administration of the aIAT to the second examinee (C.S.) revealed that average RTs (1019 ms) for the congruent block (e.g., Guilty sentence: *I killed my mother* / True) was significantly faster, *t*(119) = –9,611, *p*<.001, 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).

GENERAL DISCUSSION

The present chapter reports on a new 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 above 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, it is possible 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.

Another important issue to discuss here is that it might be said that both the aIAT and the TARA (Gregg, 2007) use response incongruity to identify lies. However, this 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 lies on the basis of an absolute level of RTs on the critical block: if the average RTs is fast then the respondent is classified as 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 2.6 of the present paper. This criminal is a medicated patient with very slow RTs. The use of the TARA in similar circumstances would require a medicated age-matched control group. Otherwise, using non-medicated controls with normal RTs 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).

A final issue is whether the aIAT can be faked. In this respect, Experiment 2.5 provides persuasive evidence that, even within an extremely faking-prone naturalistic setting, the aIAT is still able to accurately detect autobiographical event.

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).

With respect to potential countermeasures, however, some studies report that IAT (Greenwald et al., 1998) 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 that will be addressed in the next chapter.

CHAPTER 3

DETECTING FAKERS OF THE *AUTOBIOGRAPHICAL* IAT

INTRODUCTION

Lie-detectors may be used to screen suspects (e.g. terrorists at airports) or as a deterrent to reduce lapses in safety or security regimen (e.g. in nuclear plants or other defense-sensitive plants). In an investigative or forensic setting, lie-detection systems play a key role in the defense of an innocent suspect, who may accept the test in order to prove their innocence. In this specific case, faking is highly unlikely and irrational strategy. By contrast, guilty suspects have no interest in taking a test that is likely to prove their guilt. For this reason, they will be more likely either to either refuse the test or, in the remote event of accepting the test, to be prone to faking the test.

The ideal lie-detector, for investigative and forensic applications, should minimize false positive errors, which make an innocent suspect appear guilty. This error is expected when the examinee is not faking the test. By contrast, when examining a guilty suspect, who may take advantage and fake the test, false negatives, which confuse the guilty subject for an innocent subject, have to be minimized.

Effective countermeasures are now known for almost every lie-detection technique. The first study investigating the efficacy and detection of polygraphic countermeasures goes back to Benussi (1914) who also introduced the first respiratory-based lie-detection technique. Countermeasures for the CQT have long been known. Most attempts to increase the response of a subject to the control questions have been to use physical (e.g., biting the tongue or pressing the toes to the floor) or mental (e.g., counting to seven backward) techniques (Honts, Raskin, & Kircher, 1994). Countermeasures against the GKT when used with polygraph have also been demonstrated (Honts, Devitt, Winbush, & Kircher, 1996).

With respect to the aIAT countermeasures, Agosta (2005) and more recently Verschuere and colleagues (Verschuere, Prati, & De Houwer, 2009) have shown that properly trained participants may alter strategically the test outcome. Vershuere et al. (2009) instructed guilty participants in a mock-crime task to appear as innocent by slowing down their responses. Their results indicated that a large proportion of guilty participants, not previously exposed to the aIAT, succeeded in faking the test.

A critical aspect, however, which has not been fully investigated, is whether fakers could be detected on the basis of their response patterns. In fact, for the aIAT, detection of fakers should render countermeasures ineffective. A study by Fiedler and Bluemke (2005) showed that IAT experts are unable to identify faked IAT results on the basis of their expertise. Their experts were requested to evaluate the results of 24 participants (half were honest responders and half were dishonest responders) and to identify the dishonest responders. The results show that experts were unable to identify the dishonest responders on the basis of their IAT latencies. In contrast, recent evidence that IAT fakers can be detected comes from Cvencek and colleagues (Cvencek, Greenwald, Brown, Gray & Snowden, under review) who demonstrated that fakers can be uncovered by examining specific features of their response patterns.

Here, I report on a series of experiments aimed at both confirming and enhancing the validity of aIAT as a tool for evaluating autobiographical memories, even under circumstances in which faking is suspected. Although it is confirmed that the aIAT might be faked by appropriately instructing participants, here, it is also demonstrated that faking participants might be detected on the basis of their response patterns. Fakers leave a signature and, most importantly, this signature is valid for various unrelated aIATs. This signifies that this marker can be potentially used to check the authenticity of an aIAT, without a specific normative group of true and adulterated performances. Furthermore, I report on an algorithm specifically implemented for the identification of respondents who eventually succeed in faking the test.

FAKING THE *autobiographical* IAT

In this section, four experiments are described, aimed at evaluating whether participants, who were overtly instructed or simply trained previously in using an aIAT, can intentionally alter their aIAT outcome.

Methods and procedure were similar for all the experiments included in this section unless specified. The general methodology of the aIAT was the same as previously described (Chapter

2; Sartori, Agosta, Zogmaister, Ferrara, & Castiello; 2008). Here I describe the general procedure for Experiments 3.1 and 3.2 of this section.

Sentences belonging to the logical category True/False and sentences describing two autobiographical events with only one of them being true (e.g., Christmas in Paris vs. Christmas in London) were used (true and false autobiographical events were specific for each participant and collected earlier using a questionnaire). The aIAT is accomplished by requiring the respondent to complete five blocks of speeded categorization trials. Participants are requested to classify the sentences by pressing one of two labeled keys, one positioned on the left of the keyboard (e.g., "A") and the other one situated on the right of the keyboard (e.g., "L"). Sentences are presented in the center of the monitor and two reminder labels are positioned, one on the left and one on the right of the monitor. These two labels show the name of the categories that must be used in order to classify each sentence. Two out of the five blocks (critical blocks) require the double categorization of an autobiographical event (e.g., *Christmas in Paris* or *Christmas in London*) with certainly-true events (Experiment 2.1; Chapter 2).

In Block 1 (20 trials) participants had to classify certainly True or False sentences, by pressing the left key to classify certainly True sentences (five different sentences; e.g. *I am in front of a computer*) and the right key to classify certainly False sentences (five different sentences; e.g. *I am in front of a television*). In Block 2 (20 trials) participants had to classify autobiographical sentences. They pressed the left key to classify real autobiographical-event sentences (five sentences; e.g., *I saw the Eiffel Tower*) and the right key to classify false autobiographical-event sentences (five sentences; e.g., *I saw Big Ben*). In Block 3 (60 trials), the left key was used to classify both certainly True and real autobiographical event sentences, whereas the right key was used to classify both False and false autobiographical-event sentences (congruent block). In Block 4 (40 trials) the left key was used to classify false autobiographical-events sentences, whereas the right key was used to classify real autobiographical-event sentences. Finally, in Block 5 (60 trials), participants had to classify with the left key both True and false autobiographical-event sentences, and with the right key they had to classify False and real autobiographical-event sentences (incongruent block).

As the pairing of a truly autobiographical event with certainly true sentences should facilitate the response, a specific pattern of response times in the two critical blocks (3 and 5) indicates which autobiographical event is true and which autobiographical event is false.

Following an aIAT training session, participants were randomly assigned to one of three groups that differed in terms of their instructions. *Non-faking* participants received the standard aIAT instructions (i.e., they were requested to categorize the sentences as indicated by the labels by pressing the appropriate keys as fast and as accurately as possible); *naïve-faking* participants were asked to do their best to hide their true autobiographical memory to the experimenter (Fiedler & Bluemke, 2005) but they were not instructed on how to fake the test. *Instructed-faking* participants were instructed to slow down in the congruent block and speed up in the incongruent block (Kim, 2003). Note that only participants taking part in Experiment 3.1 (i.e., the Christmas aIAT experiment) were not administered the preliminary aIAT training session.

In all studies, the order of the double categorization blocks was counterbalanced across subjects (congruent block first or congruent block after the incongruent block). In the next section I describe the procedures for each of the four experiments; the findings for these experiments will be grouped and reported within the Results section.

Experiment 3.1. Faking without preliminary training with the aIAT

Forty-two participants (8 males and 34 females; age range 19-30 years) were randomly assigned to one of the three groups: 14 to the non-faking group; 14 to the naïve-faking group and 14 to the instructed-faking group. Participants were requested complete a questionnaire regarding their last Christmas holiday (e.g. *Where were you on Christmas day?*) and a Christmas holiday they never had. For each participant, a specific aIAT was built with sentences describing the true holiday and the holiday they never had. Participants pressed one of two keys corresponding to the location where they spent the holiday (e.g. *Home* (real vacation) vs. *Mountain* (false vacation). For the congruent block, true sentences and real holiday sentences were assigned to the same response key. For the incongruent block, true sentences and real holiday sentences were assigned to different keys.

Experiment 3.2: Christmas holiday aIAT with previous aIAT

In order to investigate the effect of previous aIAT experience on the ability of participants to fake the test, and on our ability to detect fakers, a second study has been run.

Fifty participants (14 males and 36 females; age range 19-30 years) took part in the experiment. Twenty participants were assigned to the non-faking group, 10 participants to the naïve-faking group, and 20 participants to the instructed-faking group. Participants were administered an aIAT pre-test session (a two-card aIAT, Experiment 2.1, Chapter 2). Then they received the Christmas aIAT as in Experiment 3.1.

Experiment 3.3: 10 cards aIAT with a preliminary training aIAT

Experiments 3.3 and 3.4 were conducted in order to generalize the results in relation to short-term memory. In these experiments participants were requested to respond to a previously-selected card and the procedure was similar to that of Experiments 3.1 and 3.2 except that sentences about the true vacation were substituted with sentences regarding the selected card and sentences regarding the false vacation were substituted with sentences regarding non-selected cards.

Seventy-two participants (20 males and 52 females; age range 19-30 years) were randomly assigned to one of the three groups: 20 participants to the non-faking group; 18 participants to the naïve-faking group and 34 participants to the instructed-faking group. At the beginning of the experiment participants were administered a preliminarily two-cards aIAT as training. Subsequently, they were requested to choose one among 10 different playing cards. After a consolidation task, consisting of identifying the previously selected card among other unrelated cards (see Experiment 2.1; Chapter 2), a subject specific “ten-cards” aIAT was administered to each participant. Here, the real autobiographical event was represented by the choice of the picked card (e.g., *I picked the card 2 of hearts*), whereas the false autobiographical event was represented by the choice of other cards (e.g. *I picked the card 3 of clubs*). One of the two reminders labels corresponded with the selected card (e.g. *2 of hearts*) whereas the other label was referred to as “other cards”.

Experiment 3.4: Two-Cards aIAT with preliminary training aIAT

Thirty-six participants (12 males and 24 females; age range 19-30 years) were randomly assigned to the three groups: 12 to the non-fakers group; 12 to the naïve-fakers group and 12 to the instructed-fakers group. Procedures and stimuli were the same as for Experiment 2.1 reported in Chapter 2, except that participants were administered a preliminary aIAT training session (a cigarette aIAT aimed at evaluating whether the respondent was a smoker or not). After the aIAT training, participants selected one of two playing cards (4 of diamonds or 7 of clubs) and were asked to memorize it during a consolidation task (see Experiment 3.3). After the consolidation task, participants performed the "two-cards" aIAT. Here, the real autobiographical event was represented by the actual choice of the card (e.g., *I picked the card number 4*), whereas the false autobiographical event was represented by the choice of the other card (e.g., *I chose the card number 7*).

RESULTS AND DISCUSSION

For all the experiments the dependent measures were RTs (between 150 and 10000 ms), D-IAT (D600 algorithm; Greenwald, Nosek, & Banaji, 2003) and accuracy. For each experiment and for each group an ANOVA on mean RTs with congruency (congruent vs. incongruent) as a within-subjects factor and order of presentation of the congruent block (congruent first in block 3 vs. congruent second in block 5) as between-subjects factor has been run. No significant main or interaction effect involving order of the presentation of the congruent block (first vs. second) emerged. Therefore only the main effect of congruency will be discussed. Table 3.1 shows mean RTs and accuracy percentage for each group.

Furthermore, an ANOVA was conducted on accuracy with congruency (congruent vs. incongruent) as a within-subjects factors and order of presentation of the congruent block (congruent first in block 3 vs. congruent second in block 5) as a between-subjects factor. No significant main or interaction effect involving order of the presentation of the congruent block (first vs. second) emerged for this second ANOVA on accuracy, except for the naïve-faking group in Experiment 3.1 ($F(1,12)= 7.594$, $p=.017$, $\eta 2 =.238$), and in Experiment 3.3 ($F(1,16)= 10.206$, $p=.006$, $\eta 2=.389$). In all experiments and in all groups accuracy was higher in the congruent than in the incongruent block, except for the instructed faker of Experiment 3.3 ($F(1,32)=20,700$, $p<.001$, $\eta 2=.393$) and 4 ($F(1,10)=0,647$, $p=.440$, $\eta 2=.061$).

EXPERIMENT 1	n. 42	Congruent	Incongruent	D-IAT	correct
Non-fakers	14	1149ms 96.90%	2104ms 84.88%	1.06	14/14
Naive-fakers	14	1360ms 95%	2045ms 92.26%	.78	14/14
Instructed-fakers	14	2381ms 98.45%	1781ms 88.21%	-0.45	5/14
EXPERIMENT 2	**n. 50**	**Congruent**	**Incongruent**	**D-IAT**	**correct**
Non-fakers	20	1163ms 95.16%	1608ms 92.16%	0.64	19/20
Naive-fakers	10	1520ms 88.5%	1692ms 83.83%	0.24	6/10
Instructed-fakers	20	1967ms 96.5%	1535ms 86.33%	-0.42	7/20

EXPERIMENT 3	n. 72	Congruent	Incongruent	D-IAT	correct
Non-fakers	20	1068ms 96.42%	1752ms 88.67%	1.13	20/20
Naive-fakers	18	1024ms 95.19%	1545ms 91.57%	0.82	18/18
Instructed-fakers	34	1976ms 82.94%	1313ms 91.76%	-0.81	4/34

EXPERIMENT 4	n. 36	Congruent	Incongruent	D-IAT (card 4)	D-IAT (card 7)	correct
Non-fakers	12	1081ms 97.77%	1317ms 99.30%	0.37	-0.52	11/12

Naive-fakers	12	1241ms 93.61%	1315ms 92.91%	0.03	0.06	6/12
Instructed-fakers	12	1155ms 93.47%	1248ms 94.44%	0.27	0.15	7/12

Table 3.1. Experiments 3.1, 3.2, 3.3 and 3.4: Summary table for the main results (RTs and accuracy). In Experiments 3.1, 3.2 and 3.3 a positive D-IAT index indicates the correct identification of the autobiographical information. Instructed-fakers are successful in trasforming the positive score into a negative score. In Experiment 3.4 positive D-IAT indicates the autobiographical information "card 4" whereas the negative D-IAT indicate "card 7". Both naïve-fakers and instructed-fakers reduce the difference between the two D-IATs when compared with non-fakers. Accuracy is generally higher for the congruent than for the incongruent block; in Experiment 3.3 only instructed-fakers do not show this pattern.

Non-Faking groups

The non-faking groups for all experiments showed faster RTs for the congruent than for the incongruent block (Experiment 3.1: RTs, *F*(1,12)=21.657, *p*<.001, η2 =.643; Experiment 3.2: *F*(1,18)=33.862, *p*<.001, η2 =.653; Experiment 3.3: *F*(1,18)=71.012, *p*<.001, η2=.798; Experiment 3.4: *F*(1,10)=11.539, *p*=.007, η2=.537).

Results for the non-faking groups (Experiments 3.1, 3.2, 3.3 and 3.4) show that the aIAT can detect the real autobiographical information with an accuracy rate above 92 %. Experiments 3.1 and 3.2 were a replication of Experiment 2.4 (Chapter 2). The only difference here was that, in Experiment 3.2, participants had previously experienced another aIAT. This previous practice did not reduce the accuracy of the aIAT in detecting the real autobiographical event (90 % in the original experiment and 95 % in this replication). However, previous practice reduced the magnitude of the D-IAT index. This is clear, when comparing non-fakers from Experiments 3.1 and 3.2, who differed only because the second group had received previous training. As shown in Figure 3.1, average D-IAT for non-fakers in Experiment 3.1 was 1.06 whereas average D-IAT for non-fakers in Experiment 3.2 was 0.64 (*F* (1,32)= 9.790, *p*= .004, η2 = .234).

The high accuracy of the procedure was also confirmed in Experiment 3.3. In this experiment participants were requested to select and memorize one of ten playing cards. In such circumstances, no classification error was observed (i.e., 100% accuracy). In Experiment 3.4, which was a replication of Experiment 2.1 (Chapter 2), accuracy reached 92%.

In sum, here it is confirmed that when participants are not instructed to alter intentionally the outcome of an aIAT, the actual autobiographical memory can be detected with high accuracy. Memory detection was equally effective for short-term memories (selected cards) and for longer term real-life memories (last Christmas vacation).

Naïve-faking groups

When participants were instructed to modify intentionally the outcome of the aIAT without receiving explicit instructions from the experimenter, the results were less consistent (Experiment 3.1: *F*(1,12)=27.423, *p*<.001, η2=.696; Experiment 3.2: *F*(1,8)=1.549, *p*=.248, η2=.162; Experiment 3.3: *F*(1,16)=29.604, *p*<.001, η2=.649; Experiment 3.4: *F*(1,10)=0.514, *p*=.490, η2=.049). In some occasions about 45% of the participants succeeded in faking (Experiment 3.2 = 40% and Experiment 3.4 = 50%). In other occasions they were unable to fake the results (Experiment 3.1 = 0% and Experiment 3.3 = 0%).

Results do not seem to be related to the preliminary training with the aIAT, given that participants did not succeed in faking the test in both Experiment 3.1 (without the preliminary aIAT), and Experiment 3.3 (with the preliminary aIAT).

Instructed-fakers

When participants were instructed explicitly to slow down on congruent trials and speed up on incongruent trials, most of them were able to modify their performance in the instructed direction (Experiment 3.1: *F*(1,12)=3.705, *p*=.078, η2=.263; Experiment 3.2: *F*(1,18)=4.583, *p*=.046, η2=.203; Experiment 3.3: *F*(1,32)=27.802, *p*<.001, η2=.465; Experiment 3.4: *F*(1,10)=0.165, *p*=.693, =.016). Therefore, participants could reverse the aIAT effect and consequently falsify the outcome of the test. The percentage of successful fakers ranges from a minimum of 42% in Experiment 3.4 to a maximum of 88% in Experiment 3.3. To sum up, I found that the aIAT is vulnerable to faking, at least when participants are explicitly instructed to slow down their responses on congruent trials. Figure 3.1 depicts the observed inversion of the D effect in this group of participants.

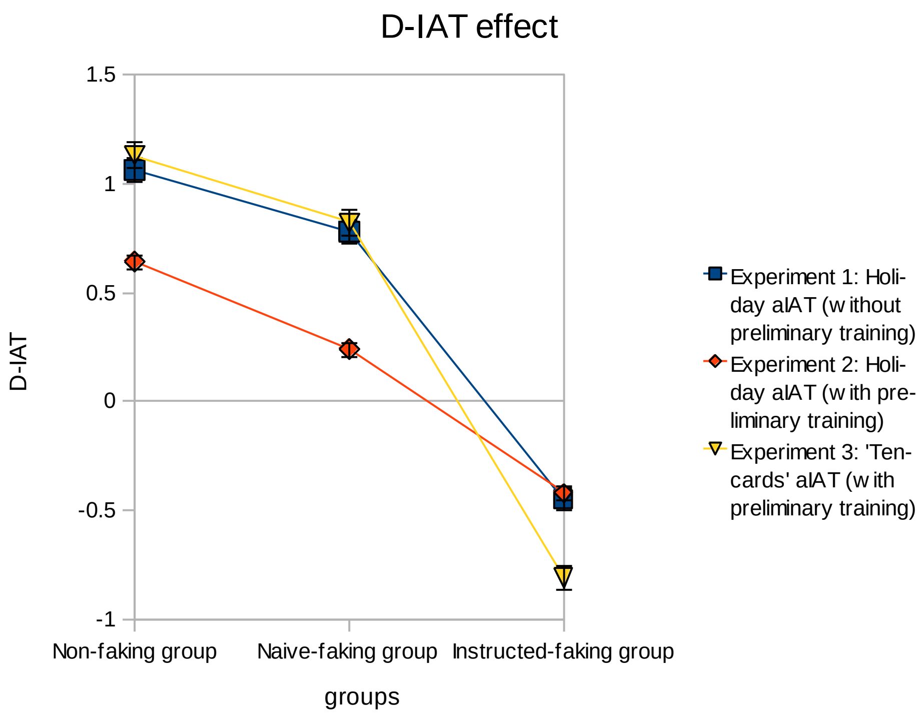


Figure 3.1. Experiments 3.1, 3.2 and 3.3: D-IAT index for the 3 groups non-fakers, naïve-fakers and instructed-fakers. The D-IAT, which is positive for the non-fakers and it indicates correct classification, becomes negative for the instructed–fakers indicating erroneous classification. As can be noticed instructed participants can revert the direction of the effect therefore succeeding in falsifying the test outcome. Naïve fakers fall somewhat in between these two groups.

DETECTING FAKERS

Findings from the series of experiments reported above show that a high proportion of participants can revert their results when instructed to do so. This result renders the aIAT vulnerable to countermeasures if used in real-life forensic settings unless an effective procedure for detecting fakers is developed.

To address this issue I further analyzed the performance of a subgroup of participants from those taking part in Experiments 3.2, 3.3 and 3.4. A total of 50 non-fakers (19 from the holiday experiment, 20 from the ten-cards experiment and 11 from the two-cards experiment) were contrasted with 58 successful fakers (17 for the Christmas experiment, 4 naïve-fakers and 13 instructed-fakers, 30 from the ten-cards experiment, 0 naïve-fakers and 30 instructed-fakers and 11 from the two-cards experiment, 6 naïve-fakers and 5 instructed-fakers). Non-fakers and fakers were selected using the D-IAT. In order to validate faking indexes which were not task dependent, relevant data from the different experiments were collapsed.

Description of the indexes

As reported above, an efficient strategy to faking effectively consists of slowing down the congruent trials. For this reason, I analyzed the difference in the RTs between the simple blocks and the double blocks in the two groups. An analysis comparing the successful fakers and correctly classified non-fakers, showing that the difference between single blocks (1, 2, and 4) and double blocks (3, 5) is larger in fakers than in non-fakers, ($F(1,106) = 26.927$, $p<.001$, η2=.203).

Therefore, here, the candidate indexes have been developed and analyzed to detect faking on the basis of this information. The two best indexes are reported in Table 3.2, in which their accuracy in classifying fakers and non-fakers is also reported. The methods that most efficiently discriminate between the two groups are variants of a comparison between RTs in double blocks (blocks 3 and 5) and single blocks (blocks 1, 2 and 4). Specifically, the efficiency of the best index was calculated using a cut-off=1.08 (Table 3.2). I considered values above the cut-off, indexing fakers, and values below the cut-off, indexing non-fakers. The cut-off was calculated using binary logistic regression under the assumption that for false alarms and missed responses had equal cost, which corresponds to the value of 50% probability of the respondent being a faker. In other words, higher values indicate that the probability of being a faker is more than 50% and lower values indicate that the probability of being a faker is fewer than 50%. Of course, in many cases costs for the two types of mistakes are different: hence, higher or lower cut-off values should be used accordingly. I also analyzed the best index (see Table 3.2) by using median RTs rather than average RTs. The median-based faking-detection index yielded an AUC (Swets, 1988) of .87 and the classification accuracy based on a Binary Logistic Regression was 81.5%.

Index	Description	AUC	Classification Accuracy using D-IAT
Ratio 150-10000 with penalty	i) Use only the RTs between 150 and 10000 ms; ii) substitute errors with the mean of the corresponding block with an added penalty of 600 ms; iii) divide the average RTs of the fastest block by the the average RTs of the corresponding single blocks (1,2 or 1,4).	0.88	82.4
Slow down 150-10000 with penalty	i) as above ; ii)as above; iii) subtract to average RTs of the fastest block (3 or 5) the average RTs of the corresponding single blocks (1,2 or 1,4).	0.88	83.3

Table 3.2. Here I report the two most efficient indexes that discriminate between fakers and non-fakers. The logic captured by these indexes is robust as I have found ten other indexes with AUC above 0.8. These non-reported indexes are variants of the two reported ones, with changes in the cut-off and penalties. The indexes encode the selective increase in RTs for the double classification blocks with respect to the single classification block that characterize successful fakers. The procedures for eliminating extremely fast and slow responses and the penalty for the errors were inspired by Greenwald (Greenwald et al., 2003). The more efficient algorithm for detecting fakers consisted of three steps: i) first eliminate all responses below 150 and above 10000 ms.; ii) substitute errors with the mean of the block added with a penalty of 600 ms iii) calculate the ratio between the average RTs of the fastest block (between 3 or 5) and single tasks that are directly connected to the fastest task in terms of motor response (1 and 2 or 1 and 4, respectively). If the result exceeds 1.08 then the respondent is faking. Threshold was identified as the cut-off which yielded the maximal classification accuracy (average index is for non-fakers 0.94 and for fakers 1.3).

The previous results were derived from participants who were analyzed in their second aIAT, after the aIAT training. In order to evaluate if the same procedure was effective in detecting fakers who did not carry out a preliminary aIAT, I analyzed a further subgroup of 14 non-fakers and 9 fakers from Experiment 3.1 (Christmas aIAT without training). In this case the classification accuracy was calculated using the index AUC of ROC analysis (AUC=.88) and 19/23 participants were correctly classified using binary logistic regression on the D-IAT. Therefore this index seems to be quite robust as it classifies very reliably participants from different aIATs, and also participants with and without previous aIAT training.

One might argue that an efficient countermeasure to faking detection might also imply a generalized slowing down for all blocks. Indeed, this strategy would invalidate the faking detection strategy, based on comparing single versus double blocks. However, this countermeasure would be quite easy to detect given that participants should manifest abnormally longer reaction times for simple blocks. By contrast, the analysis of simple blocks for all our experiments indicated that instructed fakers are not prone to slow down on single blocks but surprising they are slightly faster (Non-fakers=1184ms., sd=270; Naive fakers=1124ms., sd=414; Instructed fakers=1026 sd = 245; $F(2,188) = 5.146$, p=.007, η^2=.050).

The results reported above refer to an in-sample analysis. In order to evaluate whether the proposed detection strategy generalizes across tasks I performed a cross-validation analysis. I used data of fakers and non-fakers for Experiments 3.1 and 3.2 (i.e. autobiographical) in order to calculate the cut-off, which was then used to evaluate classification accuracy of participants (fakers and non-fakers) for Experiments 3.3 and 3.4 (i.e., cards). In-sample classification accuracy of participants for Experiments 3.1 and 3.2 was 75% (using the cut-off of 1.13) whereas out-of-sample classification of participants of Experiments 3.3 and 3.4 was 79.8% (using the same cut-off). I also calculated the cut-off using Experiments 3.3 and 3.4 (cards) and used this to classify participants for Experiments 3.1 and 3.2. In-sample classification accuracy of participants for Experiments 3.3 and 3.4 was 84.7% (cut-off 1.06) whereas out-of-sample accuracy of participants for Experiments 3.1 and 3.2 was 75.3%. These data indicate that the index used for detecting fakers may be generalized across differing tasks.

Including errors in the analysis did not improve faking detection and even if fakers produce more classification errors than non–fakers, the pattern is not very efficient in discriminating the two groups across experiments. Classification accuracy for fakers and non-fakers on the basis of the ratio between accuracy in double blocks and single blocks yielded an AUC of 0.79. These results were observed when classifying fakers and non-fakers from Experiments 3.2, 3.3, and 3.4, in which participants had a preliminary training. Accuracy analysis does not efficiently classify participants for Experiment 3.1, in which there was no preliminary training (AUC=0.46).

GENERAL DISCUSSION

The autobiographical IAT might be used as a memory-detection technique in forensic setting in which guilty suspects may be prone to faking. Four experiments, comparing non-faking participants with naïve-faking participants and instructed participants, have been conducted.

In these experiments the aIAT correctly identified the autobiographical event in non-faking participants (correct identification averaged over all experiments=96.7%). Here, it is also shown that a significant proportion of naive-fakers succeeded in faking the test using spontaneously-developed strategies (22.5%) and this proportion was much higher when participants were trained to use an optimal faking strategy (65% of them succeeded in making the experimenter believe what was not true; see also Verschuere et al., 2009).

Therefore, the aIAT could be faked using self-discovered strategies or, much more efficiently, using coached strategies. I studied the effectiveness of the self-discovered strategies of naïve-faking participants who might autonomously develop a procedure as to alter the results. In these cases, participants were instructed to fake, but they were not explicitly told how. Under these circumstances, previous experience with an aIAT facilitated the development of a self-discovered faking strategy. Indeed, when comparing the holiday aIAT without training and the holiday aIAT with training, the percentage of naïve-fakers, who succeeded in faking the test, increased from 0% to 40%.

Instructed-fakers, by contrast, were explicitly taught an optimal strategy consisting of slowing down during the congruent block and speeding up during the incongruent block. Most of them were successful in faking the test, and previous exposure to an aIAT did not increase the percentage of successful fakers. In fact, for the holiday aIAT with and without previous training, participants who succeed in faking the aIAT were 65% and 64.2%, respectively. With respect to the speeding-up observed for incongruent trials, it has been found that only in Experiment 3.1 were participants faster for incongruent trials when faking (Non-fakers=2104ms; Instructed fakers=1781ms). Speeding up for incongruent trials was not found in Experiments 3.2, 3.3, and 3.4. This difference between Experiment 3.1 and the other experiments might be due to the fact that in Experiment 3.1 a practice aIAT was not administered prior to the experimental task. Non-faking practiced participants for Experiments 3.2, 3.3 and 3.4 were presumably responding at their maximum possible speed for the incongruent trials having previously been trained with a practice aIAT.

It has not been investigated whether participants were aware of their success in faking the test, but Kim (2003) specifically tested awareness of strategies in naïve and instructed fakers of a classical attitude IAT. This author reported that only 3/24 participants, who received explicit instructions, believed they were successful in faking the test.

Noticeably, when faking, participants left behind their signature: they did not alter their RTs in single blocks and they were abnormally slow in double blocks as compared to single blocks. Here, it has been shown that this feature might be used to detect fakers with reliable accuracy. Ideally, the system for detecting fakers should generalize across subjects and conditions. I therefore tested the algorithm accuracy on participants, who were not involved in the model's development phase. Therefore, accuracy was tested with an out-of-sample procedure. Furthermore, the algorithm is equally effective for participants who did have or who did not have a preliminary aIAT. Finally, it is important to note that the algorithm does not require previous knowledge regarding the congruent block, given that this aspect would not be known in in-field applications.

The present research sought to identify an indicator of the deliberate slowing of responses that might be considered as an index of faking the aIAT outcome. The rationale underlying the development of our marker of faking was based on the observation of typical IAT results and our aIAT studies. In particular Fiedler and Bluemke (2005) showed that participants were not able to speed responses for the incongruent block. Furthermore Greenwald et al., (1998) for the IAT and Sartori et al. (2008) for the aIAT found that latencies for non-fakers in congruent blocks were comparable to those of single blocks.

In addition, Cvencek et al. (under review) reported an alternative and effective faking detection procedure which might complement the procedure presented here. These authors compared RTs for the double blocks of two IAT which were administered to the same participant. Their faking detection index, the Combined Task Slowing (CTS), consists of subtracting the difference between "the slower combined task for the faked IAT and the faster combined task for the preceding non-faked IAT". Here, in Experiments 3.2, 3.3, and 3.4, a preliminary practice aIAT (which was different from the IAT analyzed in full) has been administered to participants, and therefore it is possible to calculate the CTS on these data sets. I applied the CTS, contrasting non-fakers (n=50; belonging to the non-faking group) with participants from the instructed-faking group, with practice aIAT, who successfully faked the test (n=48). In this case, the CTS yielded an AUC equal to .75. Differences in the experimental design do not permit us to apply the index here developed to Cvencek at al.'s data. In fact, they collected responses to single blocks (which were used in our algorithm to distinguish non-fakers from fakers) only in the first non-faked IAT. In their subsequent faked IAT only double blocks were administered.

In sum, although Cvencek et al.'s faking detection index and the index reported here are based on differing logics, they both allow us to detect fakers of IAT efficiently. Indeed they might complement each other: whereas Cvencek et al.'s faking indicator might be applied when two different IATs are administered to the same subject, a first non-faked IAT and a second suspected-faked IAT, our indicator might be used when a single aIAT is administered to a suspected faker.

One might argue that the laboratory experiments reported here are very different from in-field lie detection applications in which participants might be expected to be very anxious about the results of their performance. If high anxiety is reflected itself in an increase in reaction times in double blocks, then a non-faker could, in such situations, be misclassified as a faker. In order to evaluate this hypothesis I re-analyzed the data for Experiment 2.5 which is reported in Chapter 2. Participants for the experimental group (25 participants) had their driving license suspended for driving with an excessive blood alcohol level. They were examined as part of a medico-legal screening and were let to believe that the aIAT outcome would determine whether or not their driving license would be reinstated. By contrast, control participants were never caught by the police with excessive blood alcohol level and they were tested in the laboratory. The drunk drivers group can be considered a high anxiety and low-faking group for the following reasons: i) they have no advantage in faking the test (i.e. responding as if they had not drive while drunk) because drunk driving was already established with incontrovertible evidence; ii) the setting was anxiety-prone as drunk drivers knew that the reinstatement of their driving license depended on their results of the reaction time test. Results showed that in the field anxiety did not cause a slowdown during the double blocks (average RTs on double blocks for the drunk drivers=1984ms and for the control group=1995ms), supporting the generalization of the present results to more stressful and anxious situations.

In conclusion, here it is confirmed that the aIAT is a simple, but powerful procedure for evaluating autobiographical memories. When used as a lie-detection technique it can be faked, but fakers can be identified. The indexes that have been developed are quite robust, given that minor changes in the algorithm did not cause significant reductions of classification accuracy.

Further, they provided similar classification accuracy when analyzing participants with and without preliminary IAT practice.

In sum, these results confirm for the aIAT the findings of Cvencek et al. (under review) on the IAT who concluded that *“faking of the Implicit Association Test can be detected and corrected, thus highlighting the resistance to faking as one of IATs advantages”.*

A last issue rose from Verschuere and colleagues’ paper (2008) deals with the lower accuracy in classifying non-faking participants in their experiments. In the next Chapter I will show that this lower accuracy can be explained by the use of negative sentences and labels.

CHAPTER 4

ACCURACY AND RELIABILITY OF THE *AUTOBIOGRAPHICAL* IAT

INTRODUCTION

Verschuere, Prati, & De Houwer, (2008) found, in a replication of the mock-crime experiment (Experiment 2.2, Chapter 2; Sartori et al., 2008), a lower accuracy in classifying participants than the one reported originally here and in Sartori et al., (2008) (64% vs. 91%). As regards to accuracy, I did not observe a decrease in accuracy when replicating experiments with cards and holiday aIAT (Experiment 3.1 and 3.2; Chapter 3). While the mock-crime aIAT was characterized by the use of negative reminder labels and negative sentences[2], the card and holiday aIATs used only affirmative sentences and affirmative reminder labels. This difference raises the possibility that the use of negatives in reminder labels and sentences has a detrimental effect in aIAT detection accuracy.

Here I first report on two different studies (Experiment 4.A and Experiment 4.B) aimed at evaluating the eventual detrimental effect in accuracy due to the use of negative reminder labels and negative sentences in preparing an aIAT. I will then show that using affirmative sentences results in high accuracy in identifying autobiographical events also in a replication of the mock-crime experiment (Experiment 4.C).

EXPERIMENT 4.A: 2 Cards aIAT

Here I refer to *Events* as autobiographical episodes described by a sentence (e.g. *I have been in Venice*), that can be true or false, presented always in the affirmative form; while I define a *Counter-event* as the negation of the corresponding Event (e.g. *I have not been in Venice*). Therefore, if the Event is true, the Counter-event is false, while if the Event is false, the Counter-event is true.

Participants

[2] The stimulus sentences in the aIAT, are presented on the computer screen. Reminder labels are displayed on the left and right uppermost part of the screen to facilitate recall of the meaning of the response button.

Forty students from the University of Padua volunteered for this experiment (11 males and 29 females; age range 19–30 yrs, mean age= 23.6). Out of 40 participants who took part in the study, 20 participants selected the card *4 of diamonds* and 20 participants selected the card *7 of clubs*, as described in the methods and procedures.

Methods and procedures

Two identical cards were presented face down to participants who were made to believe that the two cards were different. This procedure was used to balance the card selection. After a consolidation task, carried out to recall correctly the selected card (Experiment 2.1, Chapter 2), participants performed the experimental aIAT. Reminder labels appeared on the computer screen during the test as an aide memoire. In the original two cards aIAT the *4 of diamonds* and *7 of clubs* labels were used as reminders. By contrast, here, two types of aIAT were administered to participants. In the first aIAT, the reminder labels were *4 of diamonds-Non 4 of diamonds* (4-Non 4), whereas in the second aIAT the reminder labels were *7 of clubs-Non 7 of clubs* (7-Non 7). Sentences were affirmative if they referred to the Event (i.e. the choice of the card; e.g., *I selected card number 4*) and negative if they referred to the Counter-event (e.g., *I did not choose card number 4*).

Thus, if the Event is true (i.e. the chosen card is described by affirmative sentences; e.g. *I selected card 4*, for 4 of diamonds choosers), the Counter-event is false (i.e. the chosen card described by negative sentences; e.g. *I did not select card 4*, for card 4 of diamonds choosers). By contrast, if the Event is false (e.g. *I selected card 7*, for 4 of diamonds choosers) the Counter-event is true (e.g. *I did not select card 7*, for 4 of diamonds choosers).

The five blocks were organized as in any IAT. In Block 1 (20 trials) participants had to classify True sentences by pressing the left key and False sentences pressing the right key (five True sentences; such as: *I am in front of a computer* and five false sentences such as: *I am writing a paper*). In Block 2 (20 trials) participants had to classify sentences describing cards. They were required to press the left key to classify affirmative card sentences about the Event (five sentences such as: *I chose card 4*) and to press the right key to classify negative card sentences about the Counter-event (five sentences such as: *I did not choose card 4*). In Block 3 (60 trials), the left key was used to classify both True sentences and affirmative Event card sentences, whereas the right key was used to classify both False sentences and negative Counter-event card sentences. In Block 4 (40 trials) the left key was used to classify negative Counter-event sentences describing cards, whereas the right key was used to classify affirmative Event sentences regarding cards. Finally, in Block 5 (60 trials), participants had to classify with the left key both True sentences and negative card sentences, and with the right key they had to classify False sentences and affirmative card sentences.

As an example, I will describe in detail the 4-Non 4 aIAT. Here, if the subject chose the 4 of diamonds card than the congruent block was the block pairing True sentences with 4 of diamonds sentences and consequently False sentences and Non-4 of diamonds sentences. The incongruent block associated True sentences with Non-4 of diamonds sentences and consequently False sentences with 4 of diamonds sentences. Instead, if the subject chose the 7 of clubs card, the pattern of associations was reversed and the congruent block was the one pairing True sentences and Non-4 of diamonds sentences while the incongruent block associated True sentences with 4 of diamonds sentences.

The order of the two aIATs was counterbalanced across participants. In order 1 the congruent block was presented before the incongruent one, whereas in order 2 the sequence was reversed. Participants were assigned to one of the resulting eight experimental conditions depending on type of aIAT, order and selected card.

Results

Analyses were conducted on RTs and accuracy in Block 3 and Block 5. RTs were submitted to an analysis of variance with congruence (congruent vs. incongruent) as within subjects factor and type of aIAT (4-Non 4 vs. 7-Non7), card (4 vs. 7) and order (1 vs. 2) as between subject factors. RTs in congruent and incongruent blocks did not differ significantly (1197 ms vs. 1333 ms; $F(1,32)= 1.657$, $p=.207$, $\eta^2=.049$). The significant interaction congruence x card x type of aIAT ($F(1,32)= 17.378$, $p<.001$, $\eta^2=.352$), indicates a reversed pattern of congruent vs. incongruent blocks for the 4-Non 4 and 7-Non 7 aIATs when participants chose card 4 or card 7. In 4-Non 4 aIAT the congruent block is faster than the incongruent block only for card 4 choosers ($p=.70$), while for card 7 choosers the incongruent block is faster than the congruent one ($p=.035$). In the 7-Non 7 aIAT the congruent block is faster than the incongruent one only for card 7 choosers (*p*=.006), while for card 4 choosers the incongruent block is faster than the congruent but this difference does not reach significance (*p*=.143). Figure 4.1 shows the difference in the congruent and incongruent blocks average RTs. The ANOVA on accuracy revealed a significant interaction congruence x card x type of aIAT (*F* (1,32)= 4.389, *p*=.044, η^2=.121) and congruence x card x type of aIAT x order (*F* (1,32)= 5.603, *p*=.024, η^2=.149).

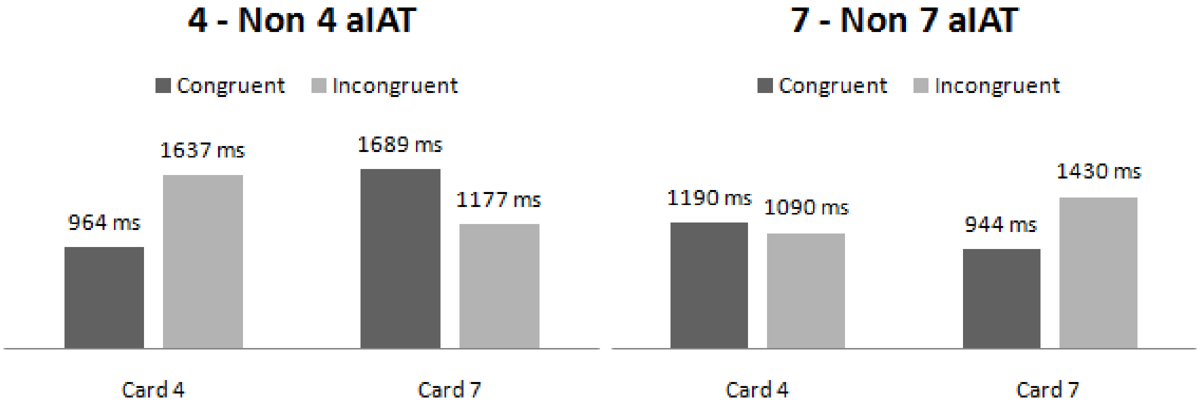


Figure 4.1: Figure 4.1 shows average RTs for 4-Non 4 aIAT and 7-Non 7 aIAT. The congruency pattern is reversed in the two aIAT depending on the chosen card. In 4-Non 4 aIAT, the congruent block for card 4 choosers (pairing affirmative card 4 sentences and True sentences) is faster than the incongruent block (pairing negative card 4 sentences and True sentences). By contrast, the incongruent block for card 7 choosers (pairing affirmative card 4 sentences and True sentences) is faster than the congruent one (pairing negative card 4 sentences and True sentences). In 7-Non 7 aIAT the congruent block for card 7 choosers (pairing affirmative card 7 sentences and True sentences) is faster than the incongruent one (pairing negative card 7 sentences and True sentences), while the incongruent block for card 4 choosers (pairing affirmative card 7 sentences and True sentences) is faster than the congruent (pairing negative card 7 sentences and True sentences).

The aIAT effect was calculated using Greenwald and colleagues' (2003) algorithm, the D-IAT expresses the difference between the two critical blocks in terms of the standard deviation of latency measures. Here I subtracted the mean of the block pairing card 4 and True sentences from the block pairing non-card 4 and True sentences in the 4-Non 4 aIAT, and the block pairing non-card 7 and True sentences from the block pairing card 7 and True sentences in the 7-Non 7 aIAT. I expected positive values for card 4 choosers and negative values for card 7 choosers. A univariate ANOVA was run with type of aIAT (4-Non 4 vs. 7-Non7) card (4 vs. 7) and order (1 vs. 2) as between subject factors and D-IAT as a dependent measure. The analysis revealed a main effect of the type of aIAT, indicating positive values for the 4-Non 4 aIAT and negatives for the 7-Non 7 aIAT (0.70 vs.-0.45; *F* (1,32)=45.634, *p*<.001, η^2=.588) and the interaction type of aIAT x order (*F* (1,32)= 10.509, *p*=.003, η^2=.247) indicating a difference in the D-IAT values in order 1 and 2 in the 4-Non 4 aIAT but not in the 7-Non 7 aIAT (*F* (1,18)= 6.238, *p*= .022, η^2=.257). If we consider the classification accuracy based on the D-IAT, when negative reminder labels and negative sentences were used (e.g., *Non 4 of diamonds* instead of *7 of clubs*; *Non 7 of clubs* instead of *4 of diamonds*) the overall accuracy of aIAT in identifying the card selected by the participants falls to 57% (in this case only 23/40 participants were classified correctly using the D-IAT; see Table 4.1). The ROC analysis revealed an AUC=.588. As already specified before, the AUC is a measure of discrimination which ranges between 1 and 0.5, in this experiment the AUC value is just above the chance level (Swets, 1988). By contrast, the same two cards aIAT but with the *Card 4-Card 7* labels yielded an overall accuracy of (95%) and a correct classification of 35/37 participants using the D-IAT (Experiment 2.1, Chapter 2; Sartori et al., 2008) with an AUC=.98.

Presented Reminder labels	**Card selected**	**Card 4**	**Card 7**	**Percentage correct classification**
4	Card 4	10	0	100%
Non 4	Card 7	8	2	20%
7	Card 7	0	10	100%
Non 7	Card 4	1	9	10%
Overall percentage				57.5%

Table 4.1. The results reported in Table 4.1 indicate that when a selected card corresponds to a category indexed by negative reminder labels and negative sentences the accuracy of aIAT is very low and below chance level indeed. By contrast, when the selected card corresponds to a category indexed by an affirmative label and affirmative sentences classification is 100%.

It is interesting to note that all the participants who selected card 4 were correctly classified, using the D-IAT, in the 4-Non 4 task. Similarly, all the participants who selected card 7 were correctly classified in the 7-Non 7 task. By contrast, participants selecting card 7 in the 4-Non 4 aIAT and those selecting card 4 in the 7-Non 7 aIAT were, in the majority, misclassified. In summary, this experiment shows that misclassifications arose for those participants who had their true memory described by the Counter-event with negative reminder labels and negative sentences.

I ran a further series of four experiments in order to verify the effects of using negative sentences and labels. In this series of experiment I contrasted affirmative and negative sentences and labels in all the four possible combinations.

EXPERIMENTS 4.B1, 4.B2, 4.B3 and 4.B4: holiday aIAT

In these experiments the effects of negative reminder labels and negative sentences on autobiographical memories (a recent holiday) are evaluated. Each experiment has been built using an autobiographical Event represented by a holiday, and a Counter-event represented by the negation of the same holiday. Two types of holiday are used: a true holiday (i.e. the last holiday that the participant had) or a false holiday (i.e. a holiday that the participant has never had).

Participants underwent, as in the previous experiment, five classification blocks. In Block 1 (20 trials) participants had to classify True sentences or False sentences. In Block 2 (20 trials) participants had to classify autobiographical sentences pressing the left key to classify Event sentences (five sentences such as: *I visited Rome*) and the right key to classify Counter-event sentences (five sentences such as: *I did not visit Rome*). In Block 3 (60 trials), the left key was used to classify both True sentences and Event sentences, whereas the right key was used to classify both False sentences and Counter-event sentences. In Block 4 (40 trials) the left key was used to classify Counter-event sentences, whereas the right key was used to classify Event sentences. Finally, in Block 5 (60 trials), participants had to classify with the left key both True sentences and Counter-event sentences, and with the right key they had to classify False sentences and Event sentences.

In the case of a true holiday, the congruent block is the one, between Block 3 and 5, pairing True sentences with Event sentences and False sentences with Counter-event sentences, while the incongruent block is the one pairing True sentences with Counter-event sentences and consequently False sentences with Event sentences. Instead, in the case of a false holiday the congruent block is the double combined block (Block 3 or 5) that pairs True sentences with Counter-event sentences (and False sentences with Event sentences); given that the holiday never took place, the real event here was represented by the negation of the false holiday, while the incongruent block was represented by the association of True sentences with Event sentences and consequently of False sentences and Counter-event sentences.

The four experiments differed in the combination of affirmative/negative sentences and labels used for the Counter-event. Sentences and labels used in the four experiments are reported in Table 4.2.

Experiment	**Event**	**Counter-event**
Experiment 4.B1	**Affirmative sentences and labels** True holiday (e.g. *I have been to Rome*), Rome False holiday (e.g. *I have been to Tokyo*), Tokyo	**Negative sentences and labels** True holiday (e.g. *I have not been to Rome*), Not Rome False holiday (e.g. *I have not been to Tokyo*), Not Tokyo
Experiment 4.B2	**Affirmative sentences and labels** True holiday (e.g. *I have been to Rome*), Rome False holiday (e.g. *I have been to Tokyo*), Tokyo	**Affirmative sentences and negative labels** True holiday (e.g. *I have been to a different place than Rome*), Not Rome False holiday (e.g. *I have been to a different place than Tokyo*), Not Tokyo
Experiment 4.B3	**Affirmative sentences and labels** True holiday (e.g. *I have been to Rome*), Rome False holiday (e.g. *I have been to Tokyo*), Tokyo	**Negative sentences and affirmative labels** True holiday (e.g. *I have not been to Rome*), Other False holiday (e.g. *I have not been to Tokyo*), Other
Experiment 4.B4	**Affirmative sentences and labels** True holiday (e.g. *I have been to Rome*), Rome False holiday (e.g. *I have been to Tokyo*), Tokyo	**Affirmative sentences and affirmative labels** True holiday (e.g. *I have been to a different place than Rome*), Other False holiday (e.g. *I have been to a different place than Tokyo*), Other

Table 4.2. The four holiday aIATs used in Experiment 4.B are summarised in Table 4.2. Here are provided examples of sentences and reminder labels for each of the aIATs used. The first comparison contrasts Events in the affirmative form (sentences and labels) with Counter-events in the negative form (sentences and labels). The second

comparison contrasts affirmative Events and Counter-events described by affirmative sentences and negative labels. The third comparison contrasts affirmative Events and Counter-events described by negative sentences and affirmative labels. The last comparison contrasts affirmative Events and affirmative Counter-events.

Experiment 4.B1: holiday aIAT (negative labels and negative sentences)

Participants

Twenty students (16 female, age range=20–44, mean age=24.5) from the University of Padua volunteered for this experiment. They were preliminarily requested to fill in a questionnaire regarding their most recent summer holidays where they were requested to briefly describe their last summer holiday and a holiday that they have never had. Then, a specific aIAT was built for each participant on the basis of the individual responses. Participants were, finally, randomly assigned to one of four conditions as described in the next section.

Method and procedure

Participants were administered one of the two types of aIAT (true holiday vs. false holiday). For the true holiday aIAT, reminder labels corresponded to the name of the location where each participant spent his actual holidays, this corresponded to the Event (e.g. reminder label=*Rome; sentence= I went to Rome last summer*). The Counter-event corresponded to the negation of the actual holiday and was represented by negative labels (e.g., reminder label= *Not Rome; sentence= I did not go to Rome last summer*). Therefore, the sentences used were affirmative if referring to the true holiday Event (e.g. *I have been to Rome*) or negative if referring to the holiday Counter-event (e.g., *I have not been to Rome*). The false holiday aIAT referred to a fictitious holiday, in this case the Event, that the participant has never had, the reminder labels corresponded to the name of a place that the participant has never visited (e.g. reminder label=*Tokyo*) and to its negation for the Counter-event (e.g., reminder label=*Non Tokyo*). The sentences were affirmative when referring to the false holiday Event (e.g., *I have been to Tokyo*) or negative when referring to the holiday Counter-event (e.g., *I have not been to Tokyo*).

Each aIAT was administered in one of two orders; in order 1 the congruent block was presented first, while in order 2 the congruent block followed the presentation of the incongruent block. Participants were assigned to one of the resulting four experimental conditions (depending on type of holiday and the order).

Experiment 4.B2 holiday aIAT (negative labels and affirmative sentences)

Participants

Twenty participants (14 females, age range 21–42, mean age=23.8) were randomly assigned to one of the four conditions as described in the next section.

Method and procedure

As in the previous experiment, participants were requested to fill in a questionnaire to describe their last holiday and a holiday that they have never had.

Participants were administered one of the two types of aIAT (true holiday vs. false holiday). The Event sentences and labels for both the true holiday and the false holiday aIAT were the same for Experiment 4.B1, while the Counter-event was represented by negative reminder labels (*Not Rome* or *Not Tokyo*) but affirmative sentences (e.g. true holiday aIAT: *I have been to a different place than Rome* vs. false holiday aIAT: *I have been to a different place than Tokyo*). The aIAT was administered in one of the two orders as described in the previous experiment.

Experiment 4.B3 holiday aIAT (affirmative labels and negative sentences)

Participants

Twenty participants (10 females, age range 19–32, mean age=23.3) were randomly assigned to one of the four conditions as described in the next section.

Method and procedure

As in the previous experiments, participants were administered one of two types of holiday aIAT (true holiday aIAT vs. false holiday aIAT). The only difference with the previous experiments is represented by sentences and labels used for the Counter-event. Labels were presented in affirmative form for both true holiday aIAT (e.g. reminder label= *Rome*) and false holiday aIAT (e.g. reminder label= *Other*), while sentences referring to Counter-events were in the negative form (e.g. true holiday aIAT: *I have not been to Rome* vs. false holiday aIAT: *I have not been to Tokyo*). Participants underwent one of the two aIAT orders.

Experiment 4.B4 holiday aIAT (affirmative labels and affirmative sentences)

Participants

Twenty participants (12 females, age range 20–38, mean age= 23.4) were randomly assigned to one of the four conditions as described in the next section.

Method and procedure

Participants were administered one of the two holiday aIATs (i.e. true holiday aIAT vs. false holiday aIAT). The Event sentences and labels were the same as described before while the Counter-event here was represented by affirmative labels (*Other* for both the true and false holiday aIAT) and sentences (e.g. true holiday aIAT: *I have been to a different place than Rome* vs. false holiday aIAT: *I have been to a different place than Tokyo*).

Result and discussion for Experiments 4.B1, 4.B2, 4.B3, 4.B4

Dependent measures were RTs between 150 ms and 10000 ms, the D-IAT (D600 algorithm; Greenwald, Nosek & Banaji, 2003) and accuracy. Reaction times and accuracy for the congruent and incongruent blocks were submitted to a mixed analysis of variance (ANOVA) with congruence (congruent vs. incongruent) as within subject factor and type of holiday (true vs. false holiday) and aIAT order (1 vs. 2) as between subject factors. The D-IAT was submitted to a univariate ANOVA with type of holiday (true vs. false) and order (1 vs. 2) as between subject factors. Here the D-IAT was calculated subtracting the average RTs in the block associating True sentences and Event sentences from the mean RTs in the block pairing True sentences and Counter-event sentences. Positive D-IAT values are expected if the Event is identified as the real fact and negative D-IAT values when the Counter-event is identified as the real fact.

With regards to RTs, the only significant result for each one of the four experiments was the interaction congruence x type of holiday (Experiment 4.B1: $F(1,16)= 30.925$, $p<.001$, $\eta^2=.659$; Experiment 4.B2: $F(1,16)= 89.389$, $p<.001$, $\eta^2=.848$; Experiment 4.B3: $F(1,16)= 45.436$, $p<.001$, $\eta^2=.740$; Experiment 4.B4: $F(1,16)= 54.255$, $p<.001$, $\eta^2=.772$). This result indicates that it is not possible to identify the real autobiographical event when this appears as the Counter-event in the false holiday aIAT. In fact, for all four experiments in the true holiday aIAT the congruent block (pairing True sentences with Event sentences) is faster than the incongruent block (pairing True sentences with Counter-event sentences), while in the false holiday aIAT the congruent block (pairing True sentences and Counter-event sentences) is slower than the incongruent block (pairing True sentences with Event sentences). Figure 4.2 shows average RTs for congruent and incongruent blocks in the four experiments.

The ANOVA on accuracy parallels the results on RTs as regards the interaction congruence x type of holiday (Experiment 4.B1: $F(1,16)= 11.835$, $p=.002$, $\eta^2=.467$; Experiment 4.B2: $F(1,16)=14.037$, $p=.007$, $\eta^2=.370$; Experiment 4.B3: $F(1,16)= 8.680$, $p=.009$, $\eta^2=.352$; Experiment 4.B4: $F(1,16)= 11.835$, $p=.003$, $\eta^2=.425$).

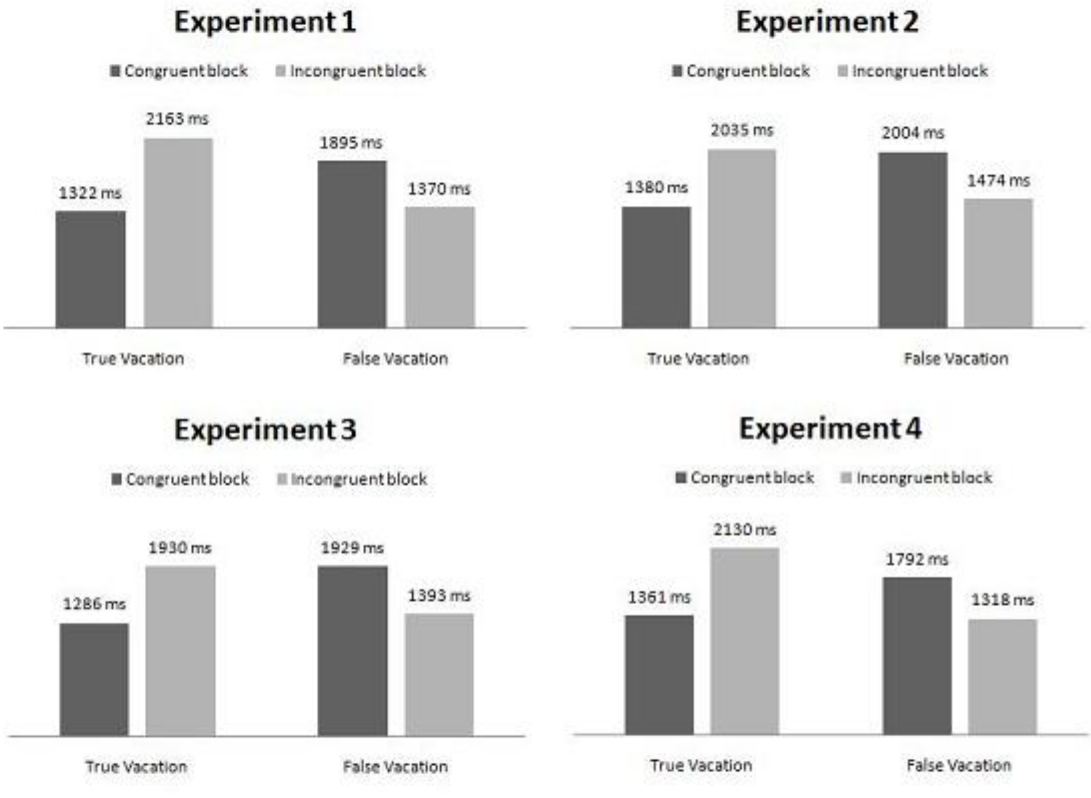


Figure 4.2: Average RTs for the congruent and incongruent blocks for Experiments 4.B1, 4.B2, 4.B3, 4.B4 are displayed in Figure 4.2. As shown here, the pattern of congruency is reversed for false holiday aIATs in respect to true holiday aIATs. In true holiday aIATs the congruent block (pairing affirmative Event sentences with True sentences) is faster than the incongruent block (pairing negative Counter-event sentences and True sentences), while in the false holiday aIAT the incongruent block (pairing Event sentences with True sentences) is faster in respect to the congruent one (pairing Counter-event sentences with True sentences).

In all four experiments the Event was correctly identified, in the true holiday aIAT, as the true autobiographical fact with an accuracy of 100%. While the Counter-event was recognized, in the false holiday aIAT, as the true fact in 10% of cases (2 out of 10) in Experiment 4.B1, in 5% of cases (1 out of 10) in Experiment 4.B2 and 0% of cases in Experiment 4.B3 and Experiment 4.B4.

In the ANOVA using the D-IAT as the dependent variable, no factor or interaction reached significance in Experiments 4.B1, 4.B2 and 4.B4. Only in Experiment 4.B3 was the interaction between type of holiday x order significant ($F(1, 16) = 6.953$, $p=0.018$, $\eta^2=.303$) indicating that for the true holiday aIAT, the D-index is higher in order 1 then in order 2 while the opposite was observed for the false holiday aIAT.

These experiments were motivated in order to improve the efficacy of the aIAT in identifying the real autobiographical episode. Experiments 4.B1 to 4.B3 showed a lower accuracy when the real autobiographical episode was presented in the form of negative sentences, negative reminder labels or both. Experiment 4.B4 showed that the accuracy of the aIAT decreases not only when using negative reminder labels and negative sentences but also when using an affirmative Counter-event. In this last experiment the accuracy for the Counter-event was 0% despite the use of affirmative reminder labels and affirmative sentences.

EXPERIMENT 4.C: Mock-crime aIAT (affirmative sentences and affirmative labels)

As mentioned before (Experiments 2.1-2.6, Chapter 2) I have used varying formats when describing false events. In two experiments, I used an affirmative format similar to the one used to describe true memories (e.g. if the true memory was the *4 of diamonds* the false memory was the *7 of clubs* etc.). I also used negative sentences for the innocent group in the mock-crime experiment (e.g. *I did not steal the CD*, Experiment 2.2, Chapter 2). Verschuere et al. (2009) replicated our mock-crime experiment and reported lower classification accuracy (Experiment 1 in Verschuere et al., 2008= 64%) than that originally reported (93%). I have shown, in the previous experiments (Experiments 2.A and 2.B, Chapter 4), that the use of negatives (as in Verschuere et al., 2009) yielded a drop in classification accuracy of experienced events that are referred by the negatives.

If the failure to replicate the high accuracy, reported by Vercheure et al. (2009), was due to the use of negative sentences then conducting the mock-crime experiment, without using negatives for refereeing innocent behavior should yield highly accurate classifications. Here two groups were contrasted: one group of "thieves" (guilty participants) that enacted a mock-crime (i.e. stealing a CD from the professor's office) and a group of "readers" (innocent participants) that read a description of the same crime in a faked newspaper article. In this case the "readers" had all the critical information as the "thieves" but did not act openly.

Participants

A total of 40 undergraduate students from the University of Padua volunteered in the study (23 females and 17 males, age range= 19–30 years old, mean age= 22.7). Half of the students were assigned to the "thieves" group (guilty group) while the other half were assigned to the "readers" group (innocent group). The "thieves" group received precise instructions to enter the professor's office and steal the CD containing the approaching written exam (Experiment 2.2, Chapter 2; Sartori et al., 2008). The "readers", by contrast, had to read a faked newspaper article reporting all the details of the event. Both groups underwent two aIATs after stealing the CD or reading the article.

Materials and methods

The aIAT consisted of the five blocks characterizing every IAT. Here the four categories and corresponding reminder labels were: the logical categories *True* vs. *False* and the autobiographical categories *Stealing* (e.g. *I stole the CD*) vs. *Reading* (e.g. *I read an article*). Reminder labels and sentences were always affirmative. In one of the two critical combined blocks participants had to classify with the same key true sentences and *stealing* sentences and consequently false sentences and *reading* sentences with the other key. In the other double categorization block they had to classify with the same response key true sentences and *reading* sentences and with the other false sentences and *stealing* sentences. Participants underwent two aIATs, one with the congruent block presented first (order 1) and the other with the incongruent block presented first (order 2). Half of the participants were administered order 1 first, while the other half were administered order 2 first.

Results and discussion

The results for the first and second aIAT administered were analyzed separately. Dependent measures were RTs between 150 and 10000 ms, the D-IAT (D600 algorithm; Greenwald, Nosek & Banaji, 2003) and accuracy. The D-IAT has been calculated as the difference between the block associating true and stealing sentences and the block associating true and reading sentences. In cases of correct classification, a positive D-IAT is expected for “thieves” while a negative D-IAT is expected for “readers”.

Reaction times and accuracy. Reaction times and accuracy for the congruent and incongruent blocks were submitted to a mixed analysis of variance (ANOVA) with congruence (congruent vs. incongruent) as within subject factor and group (thieves vs. readers) and order (1 vs. 2) as between subject factors.

Analysis of the first aIAT. Average RTs in the congruent block were significantly faster than average RTs in the incongruent block (1151 ms vs. 1448 ms; *F* (1, 36) = 49.492, *p*< 0.001, η^2=0.579). No other factors or interaction reached significance (*p*s > 0.05). The ANOVA on accuracy showed a main effect of congruence (97% accurate for the congruent block vs. 94% for the incongruent block; *F* (1, 36) = 6.225, *p*=.017, η^2=0.147).

Analysis of the second aIAT. RTs were faster in the congruent block than in the incongruent block (1071 ms vs. 1291 ms; *F* (1, 36) = 55.944, *p*<.001, η^2=.608). The significant interaction congruence x group (*F* (1, 36) = 4.598, *p*=.039, η^2=.113) indicated a larger difference between the congruent and incongruent block for the “thieves” group (283 ms) than the “readers” (157 ms). The ANOVA on accuracy showed a main effect of congruency (97% for the congruent block vs. 95% for the incongruent block; *F* (1, 36) = 8.256, *p*=.006, η^2=.188).

D-IAT. The D-IAT has been submitted to a univariate ANOVA with group (“thieves” vs. “readers”) and order (order 1 vs. order 2) as between subject factors.

Analysis of the first aIAT. “Thieves” showed, as expected, a positive D-IAT while “readers” showed a negative D-IAT (0.68 vs. -0.43; *F*(1,36)= 55.243, *p*<.001, η^2=.605). Furthermore, order 1 showed a larger D-IAT effect than order 2 (.30 vs. -.05; *F*(1,36)= 5.320, *p*=0.024, η^2=0.133). A Binary Logistic Regression analysis has been applied to the D-IAT in order to determine the accuracy of the aIAT in discriminating between “thieves” and “readers”, this analysis reported an accuracy of 87.5%. The high accuracy of the aIAT has been confirmed using the ROC analysis, the AUC for the first aIAT administered was equal to 0.94.

Analysis of the second aIAT. Thieves showed positive D-IAT while readers showed negative D-IAT (0.58 vs. -0.31; $F(1,36)= 71.142$, $p<.001$, $\eta^2=.664$). As previously underlined order 1 showed a larger D-IAT effect than order 2 (0.26 vs. 0; $F(1,36)= 5.675$, $p=.023$, $\eta^2= .136$). The Binary Logistic Regression analysis correctly classified 87.5% of the participants and the AUC was equal to .96.

I also calculated the mean of the first and second D-IAT for classifying participants. In this case, using the Binary Logistic Regression analysis it had been reached 95% accuracy and an AUC equal to 0.99.

The correlation between the D-IAT calculated on the first test and the same index calculated on the second test was r= 0.63 ($p< .001$). This index is not exactly a test-retest reliability, given that the order of the congruent and incongruent blocks were reversed in the two aIAT administrations. However, this high correlation indicates that the D-IAT is a stable measure, relatively independent from the order of presentation of the congruent block (in third position or in fifth position in the five blocks sequence).

To summarize the data collected on the mock-crime experiment, here I originally found 93% accuracy in classifying guilty and innocent participants (Experiment 2.2, Chapter 2; Sartori et al., 2008). In a replication of the same experiment, Verschuere and colleagues (2008) found, however, a lower accuracy for the same experiment (64%). This inconsistency could be due to the use of negative sentences and negative reminder labels. Here, I showed in Experiments 4.A and 4.B that the use of negative sentences and reminder labels reduces classification accuracy of individual participants when compared to a similar test using only affirmative sentences and reminder labels. For this reason I replicated the mock-crime experiment using only affirmative sentences representing two different events for both guilty (i.e. “thieves”; *I stole the CD*) and innocent (i.e. “readers”; *I read an article*) participants. In Experiment 4.C, reported here, the participants were administered two aIATs. Taken individually, both the first IAT and the second aIAT had a classification accuracy of 88%, averaging the two D-IAT the accuracy in classifying individual participants was of 95%.

GENERAL DISCUSSION

The aIAT is highly accurate in identifying an autobiographical event among two contrasting alternatives. However, there was an indication that the use of negative sentences to index autobiographical events reduces classification accuracy and leads to unreliable results.

In this chapter I report on five experiments to systematically investigate the use of negative sentences in the aIAT outcome. If affirmative sentences and reminder labels are used to describe both the true and false autobiographical events, accuracy is very high and reaches 100% in the experiments reported. By contrast, all the experiments showed that when negative sentences and negative labels are used there is a reduction in the accuracy of the aIAT in identifying the true autobiographical event. The accuracy of the aIAT is reduced, not only by negative sentences but also by affirmative sentences describing Counter-events. The affirmative Counter-event sentences were stated with expressions such as *different place than* instead of the negative (e.g. *I have been to Rome* vs. *I have been to a different place than Rome*). Negative and affirmative counter-event sentences can be considered, from this point of view, as equivalent.

One possible explanation of the detrimental effect due to the use of negative reminder labels and negative sentences refers to the figure-ground model of Rothermund & Wentura (2004). This model "assumes that the aIAT effects reflect independent salience asymmetries within the target and the attribute dimension" (Rothermund & Wentura, 2004). In brief, the two authors claim that the pattern of response is driven by the salience of the stimuli (i.e. when figures stand out from the ground) and not by the strength of the association between the two categories. Participants find it easier to respond when two salient categories are mapped onto the same motor response. This model explains the effect found by Brendl and colleagues (2001). In their experiment a strong association was present between insects and pleasant words when on the other side non-words were associated with unpleasant ones, this effect cannot be explained by implicit associations while the figure-ground model gives an explanation based on the salience of unpleasant words compared to pleasant and on the salience of non-words compared to insects (Rothermund & Wentura, 2004). In the case of our experiments, faster RTs in the block pairing true sentences and event sentences, even when the event is represented by false statements, is due to the higher salience of negative stimuli in respect to the affirmatives and by the higher salience of false stimuli in respect to the true ones (Rothermund & Wentura, 2004).

The mock-crime experiment (Experiment 4.C) shows that when we use affirmative sentences and affirmative reminder labels the accuracy of the aIAT in identifying the real autobiographical event, in the mock-crime experiment, is high. The aIAT without the use of negative reminder labels and negative sentences yielded an accuracy of 95% in classifying individual participants as guilty or innocent while in the same experiment with the use of negative reminder labels and negative sentences the corresponding reading was 64% (Verschuere et al., 2009).

The experiments reported here may be used to fine-tune guidelines for building an effective aIAT useful to identify specific autobiographical memories. As a first point, the autobiographical events that are selected for testing should be mutually exclusive (e.g. *I closed the door* vs. *I left the door open*) and maximum accuracy should be given in not including two events that are both true and false. The two events should be described in an affirmative format and referred using affirmative reminder labels. For example, suppose that we want to test whether the real autobiographical memory is about having closed a door or having left it open. In this case, sentences such as *I closed the door* should be contrasted with sentences such as *I left the door open* avoiding the words *not closed*). Adequate reminders labels could be OPEN and CLOSED. Inadequate sentences are *I did not close the door* and an inefficient reminder label is NOT OPEN.

The next step is the application of the autobiographical IAT to a new field. In all the experiments reported above, in fact, the aIAT is applied in order to identify a true autobiographical event that already happened in the past. In the next Chapter (Chapter 5) the aIAT is applied to intentions.

CHAPTER 5

INTENTION DETECTION AND ITS NEURLA BASES

INTRODUCTION

In arriving at the airport in Portland on Tuesday, September 11, 2001, Mohamed Atta had his terrorist plan already hidden somewhere in his mind. We can say that he already had the intention to hijack the American Airlines Flight 11. By contrast, the soccer world cup finalist, the French Zinedine Zidane, headbutting the Italian player Marco Materazzi, in reaction to an offence, did not have a similar planned intention.

Deliberate and explicit planning of a future action is called *prior intention* (Searle, 1983). *Prior intentions* include deliberative conscious intentions that are intuitively believed to be the leading cause of our future behaviors. In other words, these mental representations are prior to the action itself and are typically believed to subjectively *cause* the action.

Searle (1983) refers to *prior intentions* as the initial representation of the goal of an action prior to the initiation of the action; a kind of intention that is formed in advance in the form of a deliberate plan of a future action. In contrast, an *intention in action* (also defined *motor intention)* is the proximal cause of the physiological chain leading to overt behavior. When a subject has the intention (i.e. the prior intention) to perform an action, there will be many subsidiary actions that will not be represented in the prior intention. In executing a well practiced action sequence (e.g. driving the car toward the workplace), only the major headings are likely to be represented in the conscious intention (e.g., arriving at the workplace, parking, and entering the office); the details of the action sequence are unnecessary in advance (e.g., opening the car, finding a good place for parking, turning the key and opening the office). It might be said that a *prior intention* is the mental representation of the main goal of an action prior the initiation of the sequence of actions leading to the realization of the goal.

While *motor intentions* have attracted considerable attention in neuroscientific research (e.g., Libet, Gleason, Wright, & Pearl 1983; Haggard & Eimer, 1999; Lau, Rogers, Haggard & Passingham, 2004; Soon, Brass, Heinze, & Haynes, 2008), little investigation have been conducted on *prior intentions*. Typical empirical evaluations of *prior intentions* include ratings or meta-cognitive modeling such as the behavioral intention model (e.g. Fishbein & Aizen, 1975). Here prior intentions are estimated by asking subjects to rate the intensity of their intention on a rating scale. In the theory, planned behavioral intentions (which are *prior intentions*) are considered to be the results of both beliefs and subjective norms relevant for the intended behavior. This approach is grounded on a subjective evaluation of the participant about the strength of the intention itself.

Neuropsychological research focused on tasks which were aimed at evaluating a general ability called ‘prospective memory’ rather than measuring detection of specific prior intentions (Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003; McDaniel, Einstein, Graham, & Rall, 2004). Prospective memory tasks require retrieval and execution of an intention at an appropriate time or combination of circumstances, usually while another separate task is being performed. In the delayed executive paradigm (Einstein, et al., 2003; McDaniel, et al., 2004), participants are told to press a key when they encounter a specific target cue while performing an ongoing distracting task. However they were told to delay the key pressing until completing the ongoing task. On some trials, after receiving the target cue, but before completing the ongoing task, participants were interrupted by another ongoing task, which they performed until signaled to return to the original ongoing task. This form of interruption substantially impaired the delayed execution of the prospective memory task and thus is a measure of the subjects’ ability to resume the previous goal.

Few studies investigated the possibility of identifying an intention. Haynes and colleagues (2007) showed that on the basis of the activity of medial and lateral prefrontal cortex, it is possible to identify which of two tasks (adding or subtracting two numbers) the subject intends to perform in the immediate future. Here I concentrated on *what* people intend to do in the medium (i.e. during the night) and long term (i.e. their career plans) by administering to participants a modified version of the autobiographical Implicit Association Test. As in the aIAT (Chapter 2; Sartori et al., 2008), respondents are required to classify as fast as possible a series of sentences in four different categories: True (e.g. *I’m in front of a computer*), False (e.g. *I’m in front of a television*), Intention (e.g. *I will sleep in Padova tonight*), and Non-intention (e.g. *I will sleep in New York tonight*) sentences. Here the congruent condition pairs Intention sentences and True sentences while the incongruent condition pairs Non-intention sentences and True sentences. Given that the congruent condition has consistently been shown to have faster reaction times than the incongruent condition (Sartori et al., 2008), prior intentions should, therefore, be characterized by their faster RTs when paired with true sentences.

I further investigated the neural basis of the processes underlying the congruence/incongruence effect. An adapted version of the aIAT for detecting prior intentions has been administered to participants while recording their Event Related Potentials (ERPs). The aIAT effect is based on the difference, in reaction times, between an easy congruent block and a difficult, in terms of associations, incongruent block. While undergoing this last block, participants have to inhibit the tendency to make an automatic overbearing response (i.e. classify with the same motor response true sentences and true intention sentences) and select a response conflicting with it (i.e. pressing two different key for true sentences and true intention sentences). Thus, congruent and incongruent blocks are different in terms of cognitive control and conflicting responses needed to completed the tasks.

The Late Positive Component (LPC) also known as P300 is an ERP component that has been shown to be related to conflicting responses. For example, it has been shown that the LPC decreases as the conflict increases (Magliero, Bashore, Coles, & Donchin, 1984; Doucet & Stelmack, 1999) and decreases when attentional resources are located to a secondary task (Israel, Chesney, Wickens, & Donchin, 1980 a; Israel, Wickens, Chesney, & Donchin, 1980 b; Kramer, Wickens, & Donchin, 1985). Furthermore, Johnson and colleagues (Johnson, et al., 2003; Johnson, Henkell, Simon, & Zhu, 2008) showed that responses conflicting with the truth produce decreased LPC amplitudes. On the basis of these results, it is expected that LPC amplitudes would be reduced while participants undergo the incongruent task than the congruent task.

EXPERIMENT 5.1: the aIAT identifies prior intentions

Participants

A total of 22 undergraduate students of the University of Padova (6 males and 14 females; age range 18-26 years) volunteered for this study. All participants were healthy, had normal or corrected-to-normal vision, signed an informed consent and were debriefed at the end of the experiment. Participants were randomly assigned either to the Sleep group (N=11) or to the Job group (N=11). The Sleep group was tested in a medium-term prior intention (where the participant intended to sleep for the incoming night) while the Job group was tested in a long-term prior intention (the Job/profession the participant intended to do).

Materials and procedure

Sleep and Job groups were preliminarily requested to fill in a yes/no questionnaire regarding their prior intentions for the incoming night (e.g., "*Tonight I will sleep in Padua*", "*Tonight I will sleep in Milan*") and for their career planning (e.g. "*I will become a neuropsychologist*", "*I will become a lawyer*"). For each participant, an aIAT was built with Intention sentences describing either the intention for the next night (Sleep-aIAT) or the intention regarding the career planning (Job-aIAT), and Non-intention sentences describing possible intentions denied in the preliminary questionnaire. We must note that each participant was administered a different aIAT as it was built on the basis of individual intentions collected from the questionnaire. Each participant was randomly assigned to one of the two aIAT type (Sleep vs. Job). The computerized task consisted of five separate blocks of categorization trials (as illustrated in Figure 5.1).

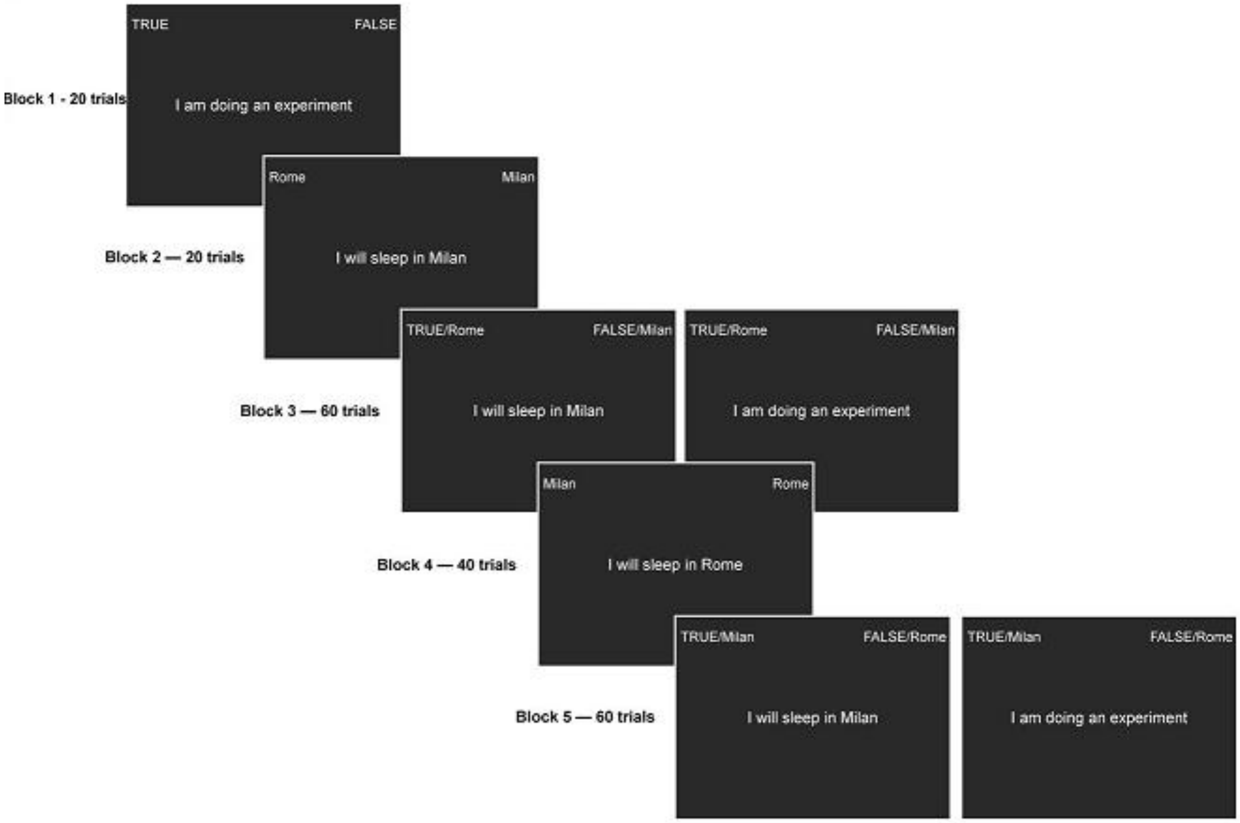


Figure 5.1: The experimental procedure of the aIAT is illustrated. Participants were required to classify the stimulus (i.e., sentences displayed) as fast as possible by pressing the correct key. In Block 3 (i.e., Congruent block) stimuli describing true actions and Intentions were assigned to the same response key (e.g., Left key). In Block 5 (Incongruent block) the same stimuli were assigned to a different response key (e.g., Right key). In block 3 and block 5 only one sentence is presented at a time.

In each trial, the stimulus sentence was presented in the center of the screen, participants were requested to classify the sentence as quickly and as accurately as possible, by pressing one of two labeled keys, one on the right (i.e., key L) and one on the left (i.e., key A) of a keyboard. In Block 1 (20 trials) participants classified stimuli along the logical dimension True vs. False by pressing left key if the sentence was True (e.g., *I am in front of a computer*) and the right key if the sentence was False (e.g., *I'm in front of a television*).

In Block 2 (20 trials) participants classified sentences along the critical dimension: True intention vs. False intention. They classified the Intention sentences (e.g., *Tonight I will sleep in Padua*) with the left key and the Non-intention sentences (e.g., *Tonight I will sleep in Milan*) with the right key. In Block 3 (60 trials, double-categorization block) they were requested to press the left key if the sentence was either True or an Intention, and the right key if the sentence was False or a Non-Intention (Congruent block). In Block 4 (40 trials) participants were requested to perform an inverted classification of the Block 2: they pressed the left key for Non-intention sentences and the right key for Intention sentences. In Block 5 (60 trials, double-categorization block) participants pressed the left key for True sentences and Non-intention sentences and the right key for False sentences and Intention sentences (Incongruent block).

Reminder labels in the form of category names were displayed on the computer screen for the entire duration of the experiment. For Intention and Non-intention categories, labels named either the location where the participant would eventually sleep the following night (e.g., Milan vs. Rome) or the career that she was planning on pursuing (e.g., Lawyer vs. Psychologist), for the Sleep aIAT and the Job aIAT, respectively. An error signal appeared for 300 ms when an incorrect response occurred.

The crucial comparison (see Figure 5.1) was between RTs in the Congruent and the Incongruent blocks. The expected pattern of implicit facilitation/inhibition should indicate participants to be faster in the block associating Intention with True sentences (congruent block), as compared to the block associating Non-intention with True sentences (incongruent block). Both for the Sleep group and the Job group, some of the participants (n=6) were presented first the congruent block (in Block 3) and then the incongruent block (in Block 5; order 1), while some (n=5; order 2) were presented the inverted order.

Results and discussion

As the previous experiments, two dependent measures were considered: mean RTs in the double-categorization blocks (3 and 5) and the D-IAT index (Greenwald, Nosek, & Banaji, 2003). RTs less than 150 ms or longer than 10.000 ms were discarded prior to any further analysis. Here the D-IAT was calculated by subtracting corrected mean RTs in the congruent block from mean RTs in the incongruent block. Then, this difference was divided by the inclusive standard deviation of the two blocks.

Data have been submitted to an analysis of variance (ANOVA) with Congruency (Congruent vs. Incongruent) as within-subjects factor and Type of aIAT (Sleep aIAT vs. Job aIAT) and Order (Order 1 vs. Order 2) as between-subjects factors. As regard to RTs, a significant

Congruency effect (F(1,18)= 70.578, p<.001, η^2=.797), indicated that average RTs for the congruent block was faster than for the incongruent block (1029 vs. 1741 ms). The significant Congruence x Type of aIAT interaction (F(1,18)= 9.340, p=.007, η^2=.342), indicated a larger aIAT effect (i.e. difference between incongruent and congruent blocks) for the Sleep group than for Job group (t(20) = -3.051, p=.006). No other interactions reached or approached significance level (all p_s > .05).

D-IAT for the Sleep group was 1.30 while for the Job group was 1.02. D-IAT was analyzed in an ANOVA with Type of aIAT (Sleep aIAT vs. Job aIAT) and Order (Order 1 vs. Order 2) as between-subjects factors. Results indicated that neither factors nor the interaction was significant (p>.05).

The Intentions of all participants were correctly detected using both RTs and D-IAT (22/22 correctly classified).

EXPERIMENT 5.2: aIAT measures prior intentions and not hopes

In Experiment 5.1, it has been found that categorization of sentences is facilitated when True sentences are paired with the same response key as Intentions. Intentions for 22/22 of the participants have been correctly identified on the basis of mean RT to the double-categorization block. However, Experiment 5.1 did not exclude the possibility that the aIAT identifies hopes (e.g. becoming a psychologist) rather than intentions.

Intentions are different from hopes; the first ones are the aims for performing actions (e.g. graduating in psychology), while the second ones reflect a belief in a positive outcome (e.g. winning the lottery). Haggard (2005) defined intentions as “several distinct processes that translate desires and goals into behavior”. Audi (1973) distinguished intentions and hopes on the basis of their probability of occurrence: “to distinguish intending to bringing about Z by doing A from merely hoping to bring about Z by doing A, we need to require that *x* (the subject) at least believes his doing A will be probable way to achieve Z”. By definition, an intention is believed to have a probable outcome, while a hope is believed to be uncertain (Audi, 1973). Both, intentions and hopes can be true but a stronger association between the “true” category and the "intention" category, in respect to the "hope" category, is expected, given that when subjects have an intention they translate this intention into behaviors, while the "hope" is believed to be uncertain.

Intentions may vary along the pleasantness of the objective to which they are associated. An objective can be highly pleasant (e.g. graduating in psychology) or not pleasant (e.g. going to the dentist), but both translate into actions aimed at the achievement of the objective (i.e. if a subject wants to graduate in Psychology, he has to pass all the exams, write a thesis etc.; if a subject wants to go the dentist, he has to call for an appointment). Thus, in Experiment 5.1 participants could have been faster in associating True sentences with Intention sentences just because Intention sentences reflect in fact a hope, instead of a real intention that will be translated into actions. For this reason, in Experiment 5.2 I directly compared prior intentions and hopes, and in particular I distinguished intentions on the basis of the pleasantness of the objective to which they are associated. I contrasted intentions associated with pleasant objectives, intentions associated with not pleasant objectives and hopes in all the three possible combinations.

Participants

Thirty students (9 male and 21 female; age range 19-30 yrs) from the University of Padova volunteered for this study. They were randomly assigned to one of three comparisons. In all the conditions the order of the double categorization blocks (congruent and incongruent block) was counterbalanced across participants. Participants were assigned randomly to one of the groups described in Table 5.1.

Materials and procedure

Prior to the experiment each participant was requested to answer a series of yes/no questions regarding 12 possible intentions (identified on the basis of the authors' intuitions). In particular they had to indicate if they had or not a series of intentions. Moreover, they had to indicate if the outcome of a specified intention was pleasant or not. An example of pleasant outcome intention was: *I will graduate in Psychology*; an example of not pleasant outcome intention was: *I will go to the dentist*. They also had to indicate if they had or not a series of hopes (an example of hope was *I will win the lottery*) and finally to indicate for each intention and hope if the probability of the outcome was high or low. They were also asked to write an example for pleasant outcome and not pleasant outcome intentions and for hopes. For each participant two categories were chosen between pleasant outcome intentions, not pleasant outcome intentions or hopes and a specific aIAT was prepared for each participant.

The aIAT procedure required the use of sentences belonging to the logical categories “true” and “false”, and sentences describing intentions differing regarding the desirability of the outcome to which they are associated or hopes. The task consisted of five classification blocks typical of any aIAT (Sartori et al., 2008). The aIAT is accomplished by requiring the respondent to complete two critical double-categorization blocks, in which intentions or hopes (e.g., “I will graduate in Psychology” or “I will go to the dentist”) are associated with certainly true events (e.g. “*I am in front of a computer*”).

In order to disentangle the critical issue whether aIAT effect can differentiate between intentions or hopes, three different comparisons were developed and three different aIAT were built. In Table 5.1 for each comparison is provided an example.

Comparison	Critical Bocks	Example	N of participants
1	Intention associated with pleasant outcome vs. Hope	Master degree vs Winnings	10
2	Intentions associated with not pleasant outcome vs. Hope	Medical Control vs. Winnings	10
3	Intention associated with pleasant outcome vs. Intention associated with not pleasant outcome	Master degree vs Medical control	10

Table 5.1. Table 5.1 shows the 3 conditions investigated in Experiment 5.2. Each condition compares intentions (related to pleasant or not pleasant outcomes) and hopes for a total of 3 comparisons. In the Table each comparison is described in terms of pleasant or not pleasant outcome associated to the intention or as a hope.

Results and Discussion

Each comparison has been analyzed separately. RTs, between 150 and 10000 ms, in the two critical double categorization blocks have been submitted to an analysis of variance (ANOVA) with double categorization blocks (type of intention, that varies for each condition, associated with True sentences) as within subject-factor and order of presentation of the double categorization blocks as between subjects factor. The order of presentation of the blocks will not be discussed further because it did not reach significance in any condition. Table 5.2 shows ANOVAs for each comparison.

In this experiment it has been shown that the aIAT identifies Intentions and not hopes. First of all I contrasted Intentions associated with pleasant outcomes and hopes, then I contrasted Intentions associated with not pleasant outcomes and hopes, and finally I compared both type of Intentions associated with pleasant and not pleasant outcomes.

In the first comparison the block pairing Intentions related pleasant outcome and true sentences is faster than the block pairing hopes and true sentences (1023 ms vs. 1518 ms, *p*<.001).

Comparisons 2 showed faster mean RTs in the block pairing Intentions related to not pleasant outcome and True sentences than the block pairing hopes and true sentences (956 ms vs. 1204 ms, *p*=.03). These two comparisons are crucial in order to disentangle the critical question if the aIAT identifies Intentions or Hopes; in particular comparison 2 showed that even when the intention is related to an unpleasant outcome, this is strongly associated with the True category.

The last comparison didn’t show any significant difference comparing both Intention related to pleasant and unpleasant outcome (1130 ms vs. 1222 ms, *p*=n.s.). This last comparison confirmed the previous results, in fact both categories are represented by Intentions and both intentions will be translated into actions.

In the first two comparisons, Intentions, either related to pleasant or unpleasant outcomes, are strongly associated with True sentences when contrasted with hope sentences, suggesting that the aIAT is identifying between Intentions and hopes the one that will be translated into actions. When comparing two intentions, even if associated with different pleasantness outcomes, I did not find any difference in terms or RTs, suggesting that both Intentions will be translated into actions.

Comparison	**Intentions with pleasant and unpleasant outcome vs Hopes**	**mean RTs (ms)**	**F (1,8)**	***p-values***	η^2	**D-IAT**
1	**Intention associated with pleasant outcome**	**1023**	**86.499**	**<.001**	**.915**	**0.92**
	Hope	1518				
2	**Intention associated with not pleasant outcome**	**956**	**6.943**	**.030**	**.465**	**0.54**
	Hope	1204				
3	**Intention associated with pleasant outcome**	**1130**	**3.206**	**Ns**		**0.24**
	Intention associated with not pleasant outcome	1222				

Table 5.2. Table 5.2 shows the ANOVA results for all the 3 conditions that were analyzed. As highlighted here, the aIAT effect is due to a difference in the put into action. Comparison 1 showed that contrasting pleasant outcome intentions and hope, only the first ones are strongly associated to the concept of truth. Comparison 2 showed the same result for not pleasant outcome intentions. Furthermore, in Comparison 3 I didn't find any difference between the two types of Intentions.

EXPERIMENT 3: The neural basis of intention detection using the aIAT

Participants

Twenty-six undergraduate students of the University of Padova (10 males and 16 females; age range 19-30) volunteered for this experiment for a credit course. They were healthy and had normal or correct-to-normal vision. All subjects signed an informed consent form and were debriefed at the end of the experiment. A participant was eliminated from analysis because he failed to comply with the instructions. Data therefore were analyzed for 25 subjects.

Procedure

The task reported in Experiment 5.1 has been adapted according to ERP requirements. Stimuli used in this experiment were twenty sentences describing intentions for the incoming night (Sleep-aIAT). True vs. False sentences and Intention vs. Non-intention sentences were all displayed one word at a time and had the same length (5 Italian words each). Participants were required to read carefully each word and to classify the sentence after the last word (i.e. *Target* word) of each sentence. Each sentence of the *true* category (e.g. "I'm in front of a computer") was matched with a sentence in the *false* category (e.g. “I'm in front of a television”) so that they differed only in the last word presented. Each sentence in the *Intention* category (“I will sleep in Padua”) was matched with a sentence in the *Non-intention* category (“I will sleep in Milan”). Also in this case they differed only in the last word. Each word was presented for 250 ms, with a 100 ms interval before the following word, with the exception of the last word of each sentence that remained on the screen as long as the subject provided the response.

In this experiment participants were administered with a greater number of trials in order to enhance the signal-to-noise ratio of ERP recording. Blocks 1 and 2 consisted of 30 trials, (each item was presented three times), Blocks 3 and 5 consisted of 120 trials (each item was presented 6 times), while the number of trials in the Block 4 was 40 trail as in the typical aIAT. Half of the participants were presented the congruent block before the incongruent one (order 1), while the other half were presented the inverted order (order 2).

EEG recording

Scalp voltages were recorded using a 59-channel electrocap with Ag/AgCl electrodes. A frontal electrode (AFz) was connected to the ground. During recording, all electrodes were referenced to mastoids. Vertical and horizontal eye movements were recorded. Electrode impedance was kept under 5 kΩ for all recordings. The EEG was recorded continuously and digitized at a sampling rate of 500 Hz. The signal was off-line filtered using a low pass filter for 30 Hz, 24 dB/octave attenuation. Ocular movements’ artifacts were corrected using the algorithm provided by Neuroscan 4.3 software. The EEG was segmented in epochs starting 200 ms before presentation of the target word and lasting until 1000 ms after its onset. The epochs were aligned to the 200 ms baseline before onset of the target word presentation. Trials contaminated by movement artifacts (peak to peak deflection over ±75µV) were rejected before averaging. The ERP were averaged for correct congruent and incongruent blocks. Approximately 5% of the trials were excluded from averaging because of movement artifacts.

Results and discussion

Behavioral data

The two dependent measures for this experiment were mean RTs (between 150 and 10000 ms) and the D-IAT (Greenwald, Nosek & Banaji, 2003). A mixed ANOVA, with Congruency (Congruent vs. Incongruent) as within subject factor and Order (Order 1 vs. Order 2) as between subject factor has been performed. RTs were measured starting from the presentation of the last word.

The D-IAT (D600 scoring algorithm, see Greenwald, Nosek & Banji, 2003) was calculated subtracting the congruent block associating True sentences and Intention sentences from the incongruent block associating True sentences and Non-Intention sentences. All individual D-IATs were submitted to an ANOVA with order (order 1 vs. order 2) as between subject factor.

As regard mean RTs, congruency factor resulted significant ($F(1,23)=8.965$, $p=.006$, $\eta^2=.280$), with response latencies in the congruent conditions were faster than in the incongruent condition (669 ms vs. 728 ms) indicating a strong association between True sentences and Intention sentences. Neither other factors nor interaction reached significance level (all $p_s > .05$).

Using D-IAT, 18/25 participants' Intentions have been correctly classified. The lower accuracy, as compared with Experiment 5.1, was presumably due to the stimuli presentation procedure used in this experiment. This lower classification accuracy replicates some previously unpublished data that I have collected using the same procedure.

ERPs results

ERP analyses were conducted only on those participants who showed a clear association between Intention sentences and true sentences. Furthermore, two participants were discarded from data analysis as they showed too many epochs with movements' artifacts. A total of 16 participants were finally analyzed.

Inspection of ERPs indicated two different components. The first was the LPC, which has been measured over center-parietal areas as in previous studies (Qu, Wang & Luo, 2008; Johnson et al., 2003, 2008). Its amplitude was determined as the mean voltage between 350 and 650 ms following the *target* word onset in congruent and incongruent blocks (Blocks 3 and 5).

The second component was the N400, a negative wave peaking in 300 and 500 ms interval (Sartori, Polezzi, Mameli, & Lombardi, 2005; Kutas & Hillyard, 1980), with maximal amplitude over parietal areas (Kutas & Federmeier, 2000). The N400 was quantified as the mean amplitude between 300 and 500 ms after the target word presentation in True and False sentences of blocks 1, 3 and 5. Because using data from multiple electrodes site may lead to a violation of the sphericity assumption, all ANOVA results were corrected using the Greenhouse-Geisser procedure.

LPC

In accordance with previous studies (Qu, Wang & Luo, 2008; Johnson et al., 2003, 2008), LPC was analyzed over midline parietal electrodes. A mixed ANOVA with Site (Cz vs. CPz vs. Pz) and Congruency (Congruent vs. Incongruent) as within subject factors, and Order (Order 1 vs. Order 2) as between subject factor was performed with LPC mean amplitude as dependent variable.

Congruency factor resulted significant ($F(1,14)=4.960$, $p<.05$, $\eta^2=.262$), with smaller LPC amplitudes for the Incongruent than for the Congruent block (2.8 vs. 3.7 µV). None of the other effects reached or approached significance level (all $p_s>.05$) (Figure 5.2). This result highlights the necessity for a greater cognitive control during the Incongruent block in respect to the Congruent one (Johnson et al., 2003).

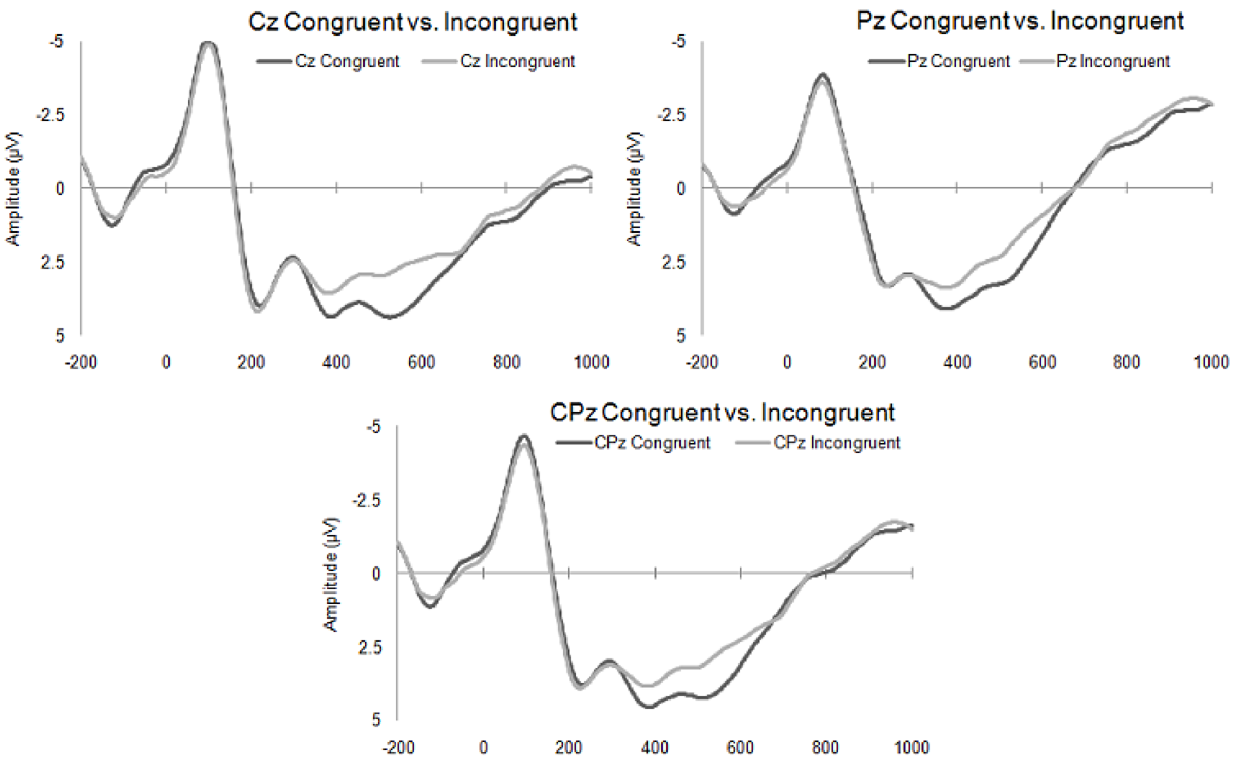
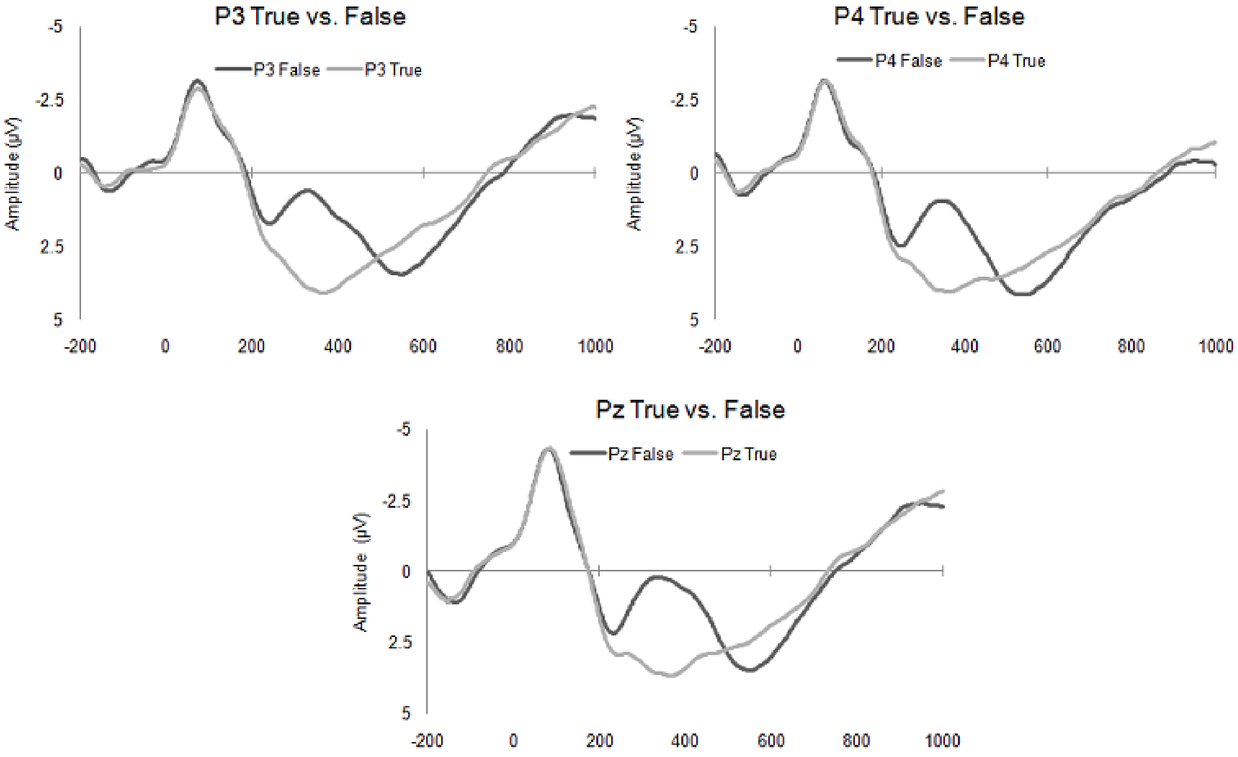


Figure 5.2. Figure 5.2 shows the Late Positive Component in Cz, CPz and Pz sites. LPC amplitudes are smaller for Incongruent than for Congruent block in all three sites. LPC is the signature con cognitive control. Smaller LPC indicates the need for a greater cognitive control.

N400

The N400 component was analyzed, for True and False sentences only, over parietal electrodes, as suggested in previous studies (Kutas & Federmaier, 2000). A mixed ANOVA with Site (P3 vs. Pz vs. P4) and Truthfulness (True vs. False) as within subject factors and order (order 1 vs. order 2) as between subject factor was performed, with N400 mean amplitude as dependent variable.

ANOVA yielded a main Truthfulness effect ($F(1,14)=20.225$, $p<.001$, $\eta^2=.591$), indicating a greater amplitude for False than for True sentences (3.5 vs. 1.4 μV). The Truthfulness x Site interaction was also significant ($F(1,28)=5.059$, $p=.013$, $\eta^2=.265$), indicating a larger difference between truth and false sentences in the central electrode compared to the right electrode ($t(15)=3.433$, $p=.004$). Figure 5.3 shows the N400 on parietal sites.


P3 True vs. False

P4 True vs. False

Pz True vs. False

Figure 5.3. Figure 5.3 depicts the N400 component in P3, Pz, P4 sites. The N400 is larger for false than for true sentences. Here the results of Hagoort and colleagues in 2004 have been replicated.

The N400 is a classical wave related to semantic congruency. For example, words out of context elicit an N400 (Kutas & Hillyard, 1980; Berkum, Hagoort & Brown, 1999). More recently, Hagoort and colleagues (2004) showed that the N400 is also elicited when a false sentence is presented, meaning that “the brain retrieves and integrates word meaning and world knowledge at the same time”.

In the present study, I found the same results by analyzing True (e.g. “I'm in front of a computer”) and False sentences (e.g. “I'm in front of a television”). It has been shown that False sentences elicit an N400 compared to True sentences.

Visual inspection of ERP comparing Intentions and Non-intention stimuli showed no difference. N400 reveals semantic incongruency and detection of a false statement (Kutas & Hillard, 1980; Hagoort et al., 2004). As Intentions and Non-intentions were not classified as true and false per se (in fact subject were instructed to classify these sentences on the basis of the type of intention not on the basis of their truthfulness), this type of items didn’t show any difference. A mixed ANOVA with Site (P3 vs. Pz vs. P4) and Type of Intention (Intention vs. Non-Intention) as within subject factors and Order (Order 1 vs. Order 2) as between subject factor was performed. No main factor or interaction reached significance ($p_s > .05$)

GENERAL DISCUSSION

The three studies reported in this chapter show that: i) it is possible to identify a *prior intention* on the basis of the aIAT congruency effect (Experiment 5.1); ii) intentions identified through the aIAT can be distinguished from hopes (Experiment 5.2); iii) the pattern of congruence/incongruence has its neural bases on the Late Positive Component (Experiment 5.3).

Here it is shown that prior intentions may be identified reliably using the aIAT. The aIAT has been primarily used to identify the true past autobiographical event (e.g. a past vacation), between two events with only one of them being true. This is the first time in which the aIAT has been used to identify a future medium and long term prior intention. Previously, Haynes and colleagues (2007) showed that it was possible to decode which of two tasks was intended to be performed in the immediate future (adding or subtracting two numbers) on the basis of a pattern of activation on the medial and lateral prefrontal activity. Here I extended the possibility of identify intentions on the basis of a pattern of congruency based on reaction time latency, to medium and long term intentions. Furthermore, I have examined whether it is possible to distinguish between Intentions or Hopes. In order to answer this critical question, I have compared pleasant and not pleasant outcome Intentions and Hopes in all of the three possible comparisons. I have provided a direct demonstration that Intentions have a stronger association with the category “True” in respect to Hopes. Both, intentions and hopes can be true but, as expected, the category that is mostly associated with the concept of truth is the one that is going to be translated into actions and behaviors (i.e. intentions).

I finally investigated the neural bases of the aIAT. The ERP analysis revealed consistently a different waveform for Congruent and Incongruent categorization. Incongruous associations showed a smaller late positivity component, compared to congruous ones in the time window between 350 and 650 ms post-stimulus.

Previous studies reported smaller LPCs when conflicting response information were introduced in the task (Magliero, et al., 1984; Doucet & Stlmack, 1999). Moreover, a series of studies demonstrated that the LPC associated to a primary-task decreased when attentional resources were allocated to a secondary task (Israel et al., 1980; Kramer et al., 1985; Wickens et al., 1983). Recent studies by Johnson and colleagues (2003; 2008) observed that responses conflicting with the truth about perceived and memorized stimuli produced a reduced LPC. In this perspective, the incongruent block, in our task, is based on inhibition of the correct answer. This answer is automatically processed and associated with the concept of “truth”, and then based on the production of the opposite response. Consequently, while a subject is engaged in the performance of the incongruent task, additional processing resources are needed. Using these extra control processes is equivalent to engaging in a *secondary task* that the person has to perform in addition to the primary task – answering the truth. All these data support the idea that the LPC may be a signature of the “response conflict” process.

I confirmed these previous results showing a smaller LPC in the incongruent block – where Intentions are associated with false sentences – than in the congruent block, where Intentions are associated with true sentences. This indicates the need of an additional control while performing the incongruent block.

CHAPTER 6

GENERAL CONCLUSIONS

PRACTICAL ASPECTS FOR THE DEVELOPMENT OF AN EFFECTIVE aIAT

Given the wide use of deception in everyday life, deception detection is drawing the attention of the scientific community. Here, I have described a new memory detection technique that has important applications in the forensic field: the autobiographical Implicit Association Test (aIAT).

The aIAT is a new technique that can be used to identify which of two autobiographical events is true. Used as a lie-detection technique the aIAT has a number of unique features when compared to traditional psychophysiological techniques of lie detection (e.g., Ben-Shakhar & Elaad, 2003) or more recent functional MRI (fMRI) based lie detection strategies (e.g., Langleben, et al., 2005). For instance, it can be administered quickly (10–15 minutes), it is based on an unmanned analysis (no training for the user is necessary), it requires low tech equipment (a standard computer is sufficient), it can be administered remotely to many participants (e.g., via the Internet) and it is possible to identify fakers on the basis of a different pattern of reaction times ratio between double blocks and single blocks (Chapter 3). The aIAT may be useful in medico-legal settings as well as in forensic sciences (Sartori, et al., 2007).

In order to build an effective aIAT it is important to follow few simple rules:

- Use short sentences that can be presented in a single line;
- Use TRUE and FALSE sentences which are always true or false for the respondent and are unrelated to the crime;
- Verify that only one of the autobiographical categories is true;
- Use reminders labels to indicate the two contrasting autobiographical information that do not include negative;
- Do not use negative or counter-event sentences;
- Before proceeding to the interpretation of the results check whether the examinee has faked or not the test by using one of the reported formulas;

- Evaluate which one of the two autobiographical sentences are associated with TRUE sentences using the D-IAT index.

LIMITS OF THE *AUTOBIOGRAPHICAL* IAT

In this last paragraph I want to highlight possible practical limits of the technique. As first issue it is important to underlie the fact that when building an aIAT it is important to have two different versions of the same event, one of which should be true. In fact the aIAT effect is maximized only if one of the two events is true and the other one is false. This issue can limit the field of applicability of the instrument to the investigative and forensic fields, where usually there are two version of the same event: one related to the prosecutor and one related to the defense.

Another important issue is related to the participant’s co-operation; in fact the test cannot be administered if the subject is not co-operative in following exactly the instructions. A related topic is the problem of the comprehension of the instructions. If participants do not comply with the instructions the test would not be valid.

Finally, I want to underline a last important aspect, as shown here, it is possible to identify fakers (on the basis of the ratio between double and single blocks), but there are not studies that investigate different types of techniques to fake the aIAT.

REFERENCES

Agosta, S. (2005). A new lie-detector based on the IAT. *Master Degree Thesis*.

Audi, R. (1973). Intending. *The Journal of Philosophy*, *70,* 387-403.

Augustine, S. (1948). “De mendacio.” In Opuscules, Vol. II, Problemes moraux. de Brouwer, Paris, pp. 244-245.

Ben-Shakar, G., (1991). Clinical judgment and decision making in CQT-polygraphy: A comparison with other pseudoscientific applications in psychology. *Integrative Psychological and Behavior Science*, *26,* 232-240.

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.

Benussi, V. (1914). Die Atmungssymptome der Lüge. *Archiv für die gesamte Psychologie, 31*, 244 - 273.

Berkum, J.J.A., Hagoort, P., & Brown, C.M. (1999). Semantic integration in sentences and discourse: evidence from the N400. Journal of Cognitive Neuroscience, *11,* 657-671.

Byrne, R.W., Machiavellian intelligence. Evolutionary Anthropology, X, 172-179.

Byrne, R.W., & Corp, N (2004). Neocortex size predicts deception rate in primates. Proceedings of the Royal Society, 271, 1693-1699.

Brendl, M., Markman, A., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in Implicit Association Test. *Journal of Personality and Social Psychology*, *81,* 760-773.

Cole, M.W. & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 1,* 343-360.

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

Cvencek, D., Greenwald, A. G., Brown, A., Gray, N. S., & Snowden, R. J. (under review). Faking of the Implicit Association Test is statistically detectable and partly correctable.

DePaulo B.M., Lindsay, J.J., Malone, B.E., Muhlenbruch, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin, 129,* 74-118.

Doucet, C., Stelmack, R.M., 1999. The effect of response execution on P3 latency, reaction time, and movement time. *Psychophysiology, 36*, 351–363.

Einstein, G.O., & McDaniel, M.A., (1990). Normal aging and prospective memory. *Journal of experimental psychology, 16*, 717-726.

Elaad, E. (2009). Effects of context and state of guilt on the detection of concealed crime. *International Journal of Psychophysiology, 71*, 225-34.

Farewell, LA, & Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, 28, 531-547.

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

Fishbein, M., Ajzen, I. (1975). *Belief, Attitude, intention, and behavior: an introduction to theory and research. Reading,* MA: Addison-Wesley.

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*, 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.

Greenwald, A.G., Mc Ghee, 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: An Improved Scoring Algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

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

Haggard, P (2005). Conscious intentions and motor cognition. *Trends in Cognitive Sciences*, *9*, 290-295.

Haggard, P. & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, *1*, 128-133.

Hagoort, P., Hald, L., Bastiaansen M., Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438-441.

Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and Physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology, 7,* 10-14.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and Physical countermeasures reduce the accuracy of the polygraph test. *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.

Isreal, J.B., Chesney, G.L.,Wickens, C.D., Donchin, F., (1980a). P300 and tracking difficulty: Evidence for multiple resources in dual performance. *Psychophysiology*, *17,* 57–70.

Isreal, J.B., Wickens, C.D., Chesney, G.L., Donchin, E., (1980b). The event-related brain potential as an index of display monitoring workload. *Human Factors*, *22*, 211–224.

Johnson, R. Jr., Barnhardt, J., & Zhu, J. (2003). The deceptive response: effects of response conflict and strategic monitoring on the Late Positive Component and episodic memory-related brain activity. *Biological Psychology, 64,* 217-253.

Johnson, R. Jr., Henkell, H., Simon, E., & Zhu, J. (2008). The self in conflict: The role of the executive processes during truthful & deceptive responses about attitudes. *Neuroimage, 39*, 469- 482.

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

Kramer, A.F., Wickens, C.D., Donchin, E., 1985. Processing of stimulus properties: evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 393–408.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207,* 203-205.

Kutas, K.D. & Federmeier (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.

LaFreniere, J.P. (1988). The ontogeny of tactical deception in humans. In Byrne, W.R. & Whiten, A. (Eds.) Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans. New York, NY, US: Clarendon Press/Oxford University Press, pp. 238-252.

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

Lau H.C., Rogers R.D., Haggard P., & Passingham R.E. (2004). Attention to intention. *Science*, *303*, 1208-1210.

Libet, B., Gleason, C.A., Wright, E.W., & Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, *106*, 623-642.

Lykken, D.T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43,* 385–388.

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

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector.* New York: Plenum Press.

Magliero, A., Bashore, T.R., Coles, M.G.H., Donchin, E., 1984. On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*, 171–186.

McDaniel, M.A., Einstein, G.O., Graham, T., & Rall, E (2004). Delaying execution of intentions: overcoming the costs of interruptions. *Applied Cognitive Psychology*, *18,* 533-547.

Mitchell, R.W. (1986). A framework for discussing deception. In Mitchell, R.W. & Thompson, N.S. (Eds.), Deception: perspectives on Human and Nonhuman Deceit. State University of New York Press, New York, pp. 3-40.

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

Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation. A technical memorandum (Report No. OTA-TM-H-15).* Washington, DC: Office of Technology Assessment.

Patrick, C. J., & Iacono, W.G. (1991). Validity of the control question polygraph test: the problem of sampling bias. *Journal of Applied Psychology, 76,* 229-238.

Qu, C., Wang, J., & Luo, Y. (2008). Inconspicuous anchoring effects generated by false information. *Progress In Natural Science*, *18,* 1375-1382.

Rosenfeld, JP, Nasman, V. T., Whalen, R., Cantwell, B. & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: a new method of lie detection. *International Journal of Neuroscience*, *34,* 125-129.

Rothermund, K., & Wentura, D (2004). Underlying processes in the Implicit Association Test: dissociating salience from associations. *Journal of Experimental Psychology*, *133*, 139-165

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., & Castiello, U. (2008). How to accurately assess autobiographical events. *Psychological Science*, *18*, 772-780.

Sartori, G., Agosta, S., & Gnoato, F. (2007). *High accuracy detection of malingered whiplash syndrome*. Paper presented at the International Whiplash Trauma Congress, Miami, FL, October 2007.

Searle J.R. (1983). Intentionality, an essay in the philosophy of mind. Press Syndacate of the University of Cambridge. Cambridge (UK)

Soon, C.S., Brass, M., Heinze, H.J., & Haynes, J.D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*, 543-545.

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.

Verschure, B., Prati, V. & De Houwer, J. (2009). Cheating the Lie Detector: Faking the Autobiographical IAT. *Psychological Science, 20*, 410-413

Wickens, C.D., Kramer, A., Vanasse, L., Donchin, F., (1983). The performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information processing resources. *Science*, *221*, 1080–1082.