Acta Psychologica 136 (2011) 269–275

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Detrimental effects of using negative sentences in the autobiographical IAT

Sara Agosta, Anna Mega, Giuseppe Sartori *

Department of Psychology, University of Padua, via Venezia 8, 35131 Padua, Italy

ARTICLE INFO

Article history:
Received 25 December 2009
Received in revised form 25 May 2010
Accepted 30 May 2010
Available online 12 January 2011

PsycINFO classification:
2340 Cognitive Processes
2343 Learning and Memory
4230 Criminal Law & Criminal Adjudication

Keywords:
Autobiographical IAT
Mock-crime
Negative sentences

ABSTRACT

The autobiographical Implicit Association Test (aIAT) is a method that accurately identifies which one of two contrasting autobiographical events is true for the subject. The aIAT indexes the real autobiographical event (e.g. *I was in Paris for Christmas*) on the basis of the facilitating effect because it maps the real autobiographical event with true sentences (e.g. *I am in front of a computer*) on the same motor response. In this paper we focus on the conditions under which the autobiographical IAT accurately and reliably identifies autobiographical memories. A recent study showed a reduction in the accuracy of the aIAT when negative sentences are used. We have investigated the detrimental effect on aIAT accuracy of such negative sentence items, used to describe autobiographical events, compared with affirmative sentence items. While we highlight the reliability of the results obtained using negative sentences, we also show that the use of affirmative sentences in describing autobiographical events guarantees high accuracy and reliability of results in identifying the true autobiographical event. Finally, we summarise the criteria for preparing stimuli for an effective aIAT in order to maximise correct classifications of individual subjects.

"Implicit attitudes are manifest as actions or judgments that are under the control of automatically activated evaluation, without the performer's awareness of that causation" (Greenwald, McGhee, & Schwarz, 1998). Implicit attitudes can be measured trough their automatic evaluation on the basis of a specific pattern of reaction times. This is the underlying mechanism at the basis of the Implicit Association Test (IAT; Greenwald et al., 1998). The idea is that it is easier and faster in terms of reaction times to map two concepts onto a single response when those concepts are somehow similar or associated in memory than when the concepts are unrelated or different. The IAT effect represents the facilitating effect due to the pairing of two associated concepts on a same response.

At present there is a debate about the origin of such implicit measures effects: whether the strength of associations (Greenwald et al., 1998) or the salience of the stimuli (Rothermund & Wentura, 2004) is responsible for the IAT effect. According to Greenwald et al. (1998), the IAT effect reflects 'mental structures involving the nominal features of IAT's categories' (i.e. names used to identify categories) and thus the strength of association between these categories (Greenwald, Nosek, Banaji, & Klauer, 2005; pg. 420), while, Rothermund and Wentura (2004) proposed that the IAT effect arises from salience asymmetries of the contrasted categories used to build an IAT. Here we report a research that cut across this debate in discussing the development of a new method for assessing the truthfulness of an autobiographical event, the autobiographical IAT (Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008).

The autobiographical IAT (aIAT; Sartori et al., 2008) is a novel variant of the Implicit Association Test (IAT) (Greenwald et al., 1998) that might be used to identify a single time-specific autobiographical event.

The aIAT, as with the standard IAT, includes stimuli belonging to four categories. Two of these categories are logical and represented by sentences which are always true (e.g. *I am in front of a computer*) or always false for the respondent (e.g. *I am in front of a television*). Two other categories are represented by autobiographical events (e.g. *I chose card number 4* vs. *I chose card number 7*), only one of the two being true. Participants have to classify sentences by pressing two response keys. The true autobiographical event, for the participant, is identified because in a combined task (when the respondent is required simultaneously to classify true and false sentences and the two autobiographical events) it gives rise to faster reaction times (RTs) when it shares the same motor response with true sentences.

Used as a lie-detection technique the aIAT has a number of unique features compared with 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 to15 min), it is based on an unmanned analysis (no training for the user is necessary), it requires low-tech equipment (a standard computer is sufficient), and it can be administered remotely to many participants (e.g. via the Internet). The aIAT may be useful in medico-legal settings as well as in forensic sciences (Sartori, Agosta, & Gnoato, 2007).

* Corresponding author. Tel.: +39 0 49 8276608; fax: +39 0 49 8276600.
E-mail address: giuseppe.sartori@unipd.it (G. Sartori).

0001-6918/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.05.011

Verschuere, Prati, and De Houwer (2009) investigated the fakeability of the aIAT and found that participants may be successfully instructed to counterfeit the aIAT outcome. The authors succeeded, through appropriate instructions, in making guilty subjects appear as innocent and vice versa. In an associated paper, we have confirmed Verschuere et al.'s (2009) results and also reported an algorithm for successfully recognising fakers of aIAT on the basis of a different pattern of reaction times ratio between double blocks and single blocks (Agosta, Ghirardi, Zogmaister, Castiello, & Sartori, 2010).

Here we focus on an important secondary result reported by Verschuere et al. (2009). The authors also studied a control group of non-fakers with the same procedure as the one reported in Sartori et al. (2008, Experiment 2). Guilty participants had to enact a mock crime and then undergo the aIAT, whereas innocent participants had to read an article reporting the same crime before being tested with the aIAT. This control group of innocents was, therefore, a replication of the original mock-crime experiment reported in Sartori et al. (2008), except for the irrelevant aspect of the language spoken by the respondents.

Verschuere et al. (2009) reported a lower accuracy in classifying the participants than that originally observed by Sartori et al. (2008; 64% vs. 93%). By contrast, other replications of the aIAT, with other experiments, confirmed the original figures and, given the importance of the issue, we decided to analyse further the origin of the reduced accuracy reported by Verschuere et al. (2009).

Successful replications included the card experiment (original: Sartori et al. (2008), accuracy = 92%; replication: Agosta et al. (2010), accuracy = 91.5%). In this experiment participants had to choose one of two cards that were held face down on a table. After choosing the card, they were administered the aIAT to identify which one of the two cards was selected by the participant. Other successful replications included the holiday experiment (original: Sartori et al. (2008), accuracy = 91%, replication: Agosta et al. (2010), accuracy = 92%). Here participants had to complete a questionnaire regarding their last holiday and one that they had never had, then they were administered the aIAT to identify the real holiday.

In order to evaluate the origin of the 'failure-to-replicate' reported by Verschuere et al. (2009) we focused on differences between the characteristics of the replicated experiments (cards and autobiographical) and the non-replicated one (mock crime). One major difference was the following: whereas the mock-crime aIAT was characterised by the use of negative reminder labels and negative sentences[1] for one of the two events, the card and holiday aIATs used only affirmative sentences and affirmative reminder labels for both events. This difference raises the possibility that the use of negatives in reminder labels and sentences has a detrimental effect on aIAT detection accuracy.

To analyse this possibility further, we first report on two different studies (Experiment A and Experiment B) aimed at evaluating the potential detrimental effect in accuracy owing to the use of negative reminder labels and negative sentences in preparing an aIAT. We will then show that using affirmative sentences also results in high accuracy in identifying autobiographical events in a replication of the mock-crime experiment (Experiment C).

We anticipate that the investigation into the use of negative sentences in describing autobiographical memories will lead to the conclusion that negative sentences is a major cause in the misdiagnosis of participants.

[1] The stimulus sentences in the aIAT are presented on the centre of 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.

1. Experiment A: card aIAT

Throughout the paper we refer to Events as autobiographical episodes described by a sentence (e.g. *I have been in Venice*) that can be true or false, presented every time in the affirmative form; we 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, whereas if the Event is false, the Counter-event is true.

Experiment A was run in order to compare directly the procedure that we used in the original card experiment reported by Sartori et al. (2008), where two affirmative Events were contrasted (i.e. the choice of card 4 or the choice of card 7), and a procedure contrasting an Event and a Counter-event (i.e. the choice of card 4 or the non-choice of card 4). Here participants, after choosing one of two cards, were administered a card aIAT: the Event was represented by the choice of a card whereas the Counter-event was represented by the negation of this choice.

1.1. Participants

Forty students from the University of Padua volunteered for this experiment (11 males and 29 females; age range 19 to 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.

1.2. Methods and procedure

Two identical cards were presented face down to participants, who were led 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 (Sartori et al., 2008), participants performed the experimental aIAT. Reminder labels appear on the computer screen during the test as an aide memoire. In the original two-card aIAT the *4 of diamonds* and *7 of clubs* labels were used as reminders. In 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 was true (i.e. the chosen card was described by affirmative sentences; e.g. *I selected card 4*, for 4 of diamonds choosers), the Counter-event was false (i.e. the chosen card was described by negative sentences; e.g. *I did not select card 4*, for 4 of diamonds choosers). By contrast, if the Event was false (e.g. *I selected card 7*, for 4 of diamonds choosers) the Counter-event was true (e.g. *I did not select card 7*, for 4 of diamonds choosers).

The five blocks were organised as in any IAT. In Block 1 (20 trials) participants had to classify True sentences by pressing the left key and False sentences by pressing the right key (there were 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 cards with negative Counter-event sentence

cards, whereas the right key was used to classify cards with affirmative Event sentences s. 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, we will describe in detail the 4–Non 4 aIAT. Here, if the subject chose the 4 of diamonds card then 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. Conversely, 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 whereas 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.

1.3. Results and discussion

Analyses were conducted on reaction times and errors in the congruent and incongruent blocks and also on D-IAT, a comprehensive measure that takes into account both latencies and errors and that will be discussed later in details. As the analyses on RTs and accuracy parallel those on the D-IAT they will not be reported in details. Fig. 1 reports average RTs.

The aIAT effect was calculated with Greenwald et al.'s (2003) algorithm; the D-IAT expresses the difference between the two critical blocks in terms of the standard deviation of latency measures. The D-IAT is the standardized measure to be used when classifying subjects; in fact it measures the difference between the congruent and the incongruent block for each subject, thus providing an index of the difference of association between the two Events and the logical category True.

Here we 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. We 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; $F(1,32)=45.634$, $p<0.001$, $^2=0.588$) and the interaction type of aIAT x order ($F(1,32)=10.509$, $p=0.003$, $^2=0.247$), 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 ($F(1,18)=6.238$, $p=0.022$, $^2=0.257$).

Table 1 shows hit rates, based on D-IAT, for this and the following experiments, comparing the classification accuracy with that reported originally in Sartori et al.'s (2008) experiment and Verschuere et al.'s (2009) experiment. Classification accuracy of the negative card experiment was 57.5%. By contrast, the same two cards aIAT experiment but with *Card 4–Card 7* labels yielded an overall accuracy 95% and a correct classification of 35/37 participants using the D-IAT (Sartori et al., 2008). It is interesting to note that all the participants who selected card 4 were correctly classified by 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 largely 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).

We ran a further series of four experiments in order to verify the effects of using negative sentences and labels. In this series of experiments we contrasted affirmative and negative sentences and labels in all the four possible combinations on a different type of autobiographical event.

2. Experiment B: holiday aIAT

In this experiment we evaluated the effects of negative reminder labels and negative sentences on autobiographical memories of a recent holiday. The experiment was built from an autobiographical Event represented by a holiday and a Counter-event represented by the negation of the same holiday. Two types of holiday were used: a true holiday (i.e. the last holiday that the participant had) or a false holiday (i.e. a holiday that the participant had never had).

Participants responded, as in the previous experiment, to 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 by 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

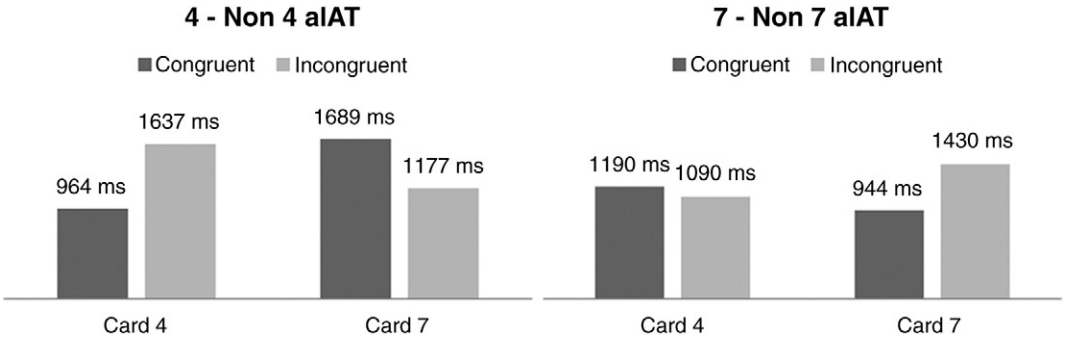

Fig. 1. This figure shows average RTs for 4–Non 4 aIAT and 7–Non 7 aIAT. The congruency pattern is reversed in the two aIAT depending on the chosen card. In 4–Non 4 aIAT, the congruent block for card 4 choosers (pairing affirmative card 4 sentences and True sentences) is faster than the incongruent block (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), whereas 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).

Table 1
This table shows hit rates for the three experiments reported here compared with hit rates from previous experiments in Sartori et al. (2008) and Verschuere et al. (2009). The use of negative reminder labels and negative sentences reduces the hit rate compared with the use of affirmative labels and sentences in all three experiments reported here. When two groups are compared, the accuracy percentage is the result of the Binary Logistic Regression analysis that correctly predicts the category of outcome for individuals on the assumption that false alarms and missed responses have equal costs; when two events are compared, however, the accuracy is the result of the D-IAT classification using zero as the cut-off value.

Experiment	Hit rate	Affirmative experiment	Negative experiment
Card aIAT	Classification accuracy	94.6% (Exp.1; Sartori et al., 2008)	57% (Exp. A)
Holiday aIAT	Classification accuracy	90% (Exp. 4; Sartori et al., 2008)	51.8% (Exp. B)
Mock-crime aIAT	Classification accuracy	95% (Exp. C)	93% (Exp. 2; Sartori et al., 2008) 64% (Exp. 1, controls; Verschuere et al., 2009)

Table 2
The four holiday aIATs used in Experiment B are summarised in Table 2, which provides 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	Event	Counter-event
Experiment B1	Affirmative sentences and labels	Negative sentences and labels
	True holiday (e.g. *I have been to Rome*), Rome	True holiday (e.g. *I have not been to Rome*), Not Rome
	False holiday (e.g. *I have been to Tokyo*), Tokyo	False holiday (e.g. *I have not been to Tokyo*), Not Tokyo
Experiment B2	Affirmative sentences and labels	Affirmative sentences and negative labels
	True holiday (e.g. *I have been to Rome*), Rome	True holiday (e.g. *I have been to a different place than Rome*), Not Rome
	False holiday (e.g. *I have been to Tokyo*), Tokyo	False holiday (e.g. *I have been to a different place than Tokyo*), Not Tokyo
Experiment B3	Affirmative sentences and labels	Negative sentences and affirmative labels
	True holiday (e.g. *I have been to Rome*), Rome	True holiday (e.g. *I have not been to Rome*), Other
	False holiday (e.g. *I have been to Tokyo*), Tokyo	False holiday (e.g. *I have not been to Tokyo*), Other
Experiment B4	Affirmative sentences and labels	Affirmative sentences and affirmative labels
	True holiday (e.g. *I have been to Rome*), Rome	True holiday (e.g. *I have been to a different place from Rome*), Other
	False holiday (e.g. *I have been to Tokyo*), Tokyo	False holiday (e.g. *I have been to a different place from Tokyo*), Other

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 Blocks 3 and 5, pairing True sentences with Event sentences and False sentences with Counter-event sentences, whereas the incongruent block is the one pairing True sentences with Counter-event sentences and consequently False sentences with Event sentences. Conversely, in the case of a false holiday the congruent block is the one 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 is represented by the negation of the false holiday, whereas the incongruent block is represented by the association of True sentences with Event sentences and consequently that of False sentences and Counter-event sentences.

2.1. Participants

Eighty students (52 female, age range = 19–44, mean age = 23.75) from the University of Padua volunteered for this experiment. They were initially requested to fill in a questionnaire regarding their most recent summer holidays: they were requested to describe their last summer holiday briefly and a holiday that they had 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 sixteen conditions, as described in the next section.

2.2. Material and procedure

Four different conditions were included in the combination of affirmative/negative sentences and labels used to index the Counter-event. Sentences and labels used in the four conditions are reported in Table 2.

Participants were administered one of the two types of aIAT (true holiday vs. false holiday) and one of two orders (in order 1 the congruent block was presented first, whereas in order 2 the congruent block followed the presentation of the incongruent block). Each participant was then administered one of four versions of B experiment (B1, B2, B3, and B4) differing in the use of negative/positive labels and sentences for the Counter-event. Participants were assigned to one of the resulting sixteen experimental conditions (depending on type of holiday, order and version of the experiment).

Experiment B1 (negative labels and negative sentences): for the true holiday aIAT, reminder labels for the Event corresponded to the name of the location where each participant spent her/his actual holidays (e.g. reminder label = *Rome*). The Counter-event corresponded to the negation of the actual holiday and was represented by negative labels (e.g. reminder label = *Not Rome*). 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 had never had; the reminder labels corresponded to the name of a place that the participant had 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*).

Experiment B2 (negative labels and positive sentences): the Event sentences and labels for both the true holiday and the false holiday aIAT were the same for Experiment B1, whereas 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 from Rome* vs. false holiday aIAT: *I have been to a different place from Tokyo*).

Experiment B3 (affirmative labels and negative sentences): the only difference between this and 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*), whereas 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 B4 (affirmative labels and affirmative sentences): the Event sentences and labels were the same as described before and 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 from Rome* vs. false holiday aIAT: *I have been to a different place from Tokyo*).

2.3. Results and discussion

Analyses were conducted on RTs (between 150 and 10,000 ms) and the D-IAT (D600 algorithm; Greenwald, Nosek, & Banaji, 2003). Also here, as before, we will report only analyses for the D-IAT as these results are comparable to those obtained on raw RTs. Fig. 2 shows average RTs for congruent and incongruent blocks separately for the four versions of the experiment.

The D-IAT was submitted to a univariate ANOVA with version of the experiment (B1, B2, B3, and B4), type of holiday (true vs. false) and order (1 vs. 2) as between-subject factors. Here the D-IAT was calculated by 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. In the ANOVA using the D-IAT as the dependent variable only the factor type of holiday reached significance ($F(1,64) = 6.875$, $p = 0.011$, $^2 = 0.097$), indicating a greater D-IAT value for the true vacation aIAT than for the false vacation aIAT. 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 versions of Experiment B 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), whereas 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).

Table 1 shows hit rates for the false holiday experiment (51.8%) compared with results from the original holiday experiment, where two Events were contrasted (90%; Experiment 4, Sartori et al., 2008). In all four versions of Experiment B the Event was correctly identified in the true holiday aIAT as the true autobiographical fact with an accuracy of 100%. By contrast, the Counter-event was recognised in the false holiday aIAT as the true fact in 10% of cases (2 out of 10) in Experiment B1, in 5% of cases (1 out of 10) in Experiment B3 and 0% of cases in Experiment B2 and Experiment B4.

Experiments B1 to B3 showed lower classification accuracy compared with the original autobiographical aIAT experiment presented in Sartori et al. (Experiment 4; 2008) when the real autobiographical episode was presented in the form of negative sentences, negative reminder labels or both. Experiment B4 showed that the accuracy of the aIAT decreases not only when negative reminder labels and negative sentences are used but also when an affirmative Counter-event is used. In this case the accuracy for the Counter-event was 0% despite the use of affirmative reminder labels and affirmative sentences.

3. Experiment C: mock-crime aIAT (affirmative sentences and affirmative labels)

As mentioned before, Sartori et al. (2008) used different verbal formats when describing false events. In two experiments, we 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.). In another case we used negative sentences to index innocent behaviour in the mock-crime experiment (e.g. *I did not steal the CD*, Experiment 2, Sartori et al., 2008). 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%). We have shown, in the previous experiments (Experiments A and B), that the use of negatives (as in Verschuere et al., 2009; Sartori et al., 2008) may yield a drop in classification accuracy of experienced events that are addresses by the negatives.

If the lower classification accuracy, reported by Verschuere et al. (2009), resulted from the use of negative sentences using only

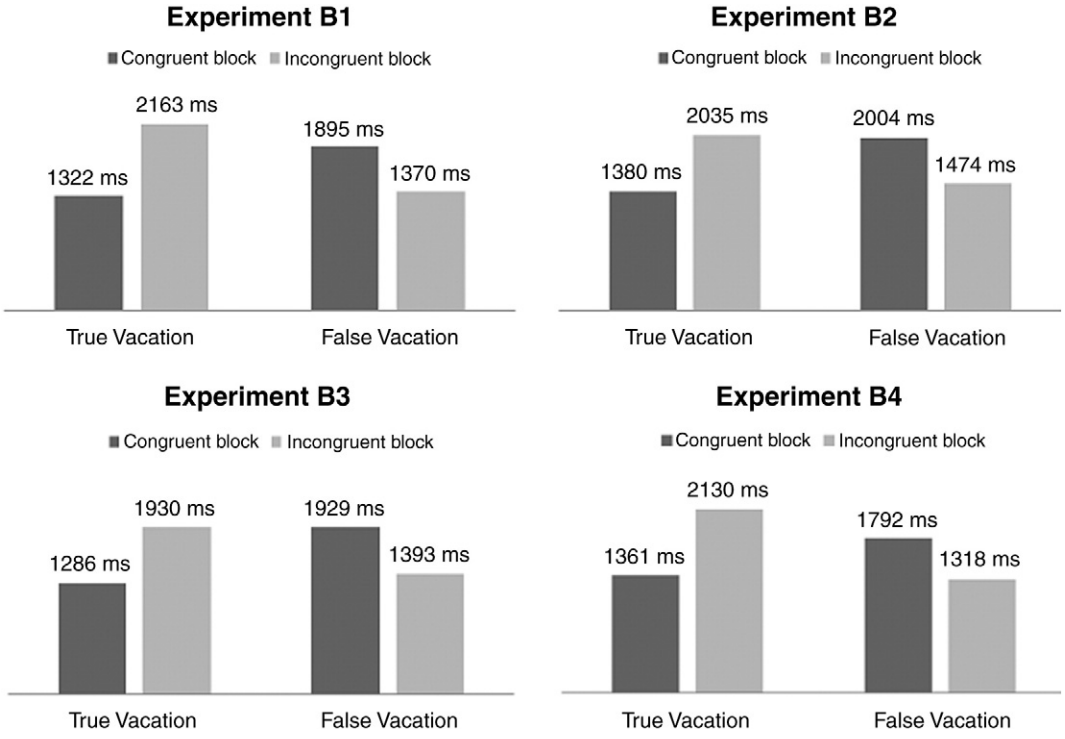

Fig. 2. Average RTs for the congruent and incongruent blocks for Experiments B1, B2, B3, and B4 are displayed in this figure. As shown here, the pattern of congruency is reversed for false holiday aIATs in respect of true holiday aIATs. In true holiday aIATs the congruent block (pairing affirmative Event sentences with True sentences) is faster than the incongruent block (pairing negative Counter-event sentences and True sentences), whereas in the false holiday aIAT the incongruent block (pairing Event sentences with True sentences) is faster in respect of the congruent one (pairing Counter-event sentences with True sentences).

affirmative sentences for referring to innocent behaviour should yield higher accuracy. In order to compare directly the use of negative and positive sentences and labels to describe the same event for the mock-crime experiment we ran Experiment C. Here two groups were contrasted: one group of 'thieves' (guilty participants) who enacted a mock crime (i.e. stealing a CD from the professor's office) and a group of 'readers' (innocent participants) who read a description of the same crime in a faked newspaper article. In this case the 'readers' had the same critical information as the 'thieves' but did not enact the mock crime.

3.1. Participants

A total of 40 undergraduate students from the University of Padua volunteered for the study (23 females and 17 males, age range = 19 to 30 years old, mean age = 22.7). Half of the students were assigned to the 'thieves' group (guilty group) and 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 an exam paper (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.

3.2. Materials and methods

The aIAT consisted of the five blocks characterising 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 combined blocks, participants had to classify with the same key true sentences and *stealing* sentences and subsequently false sentences and *reading* sentences with the other key. In the other double categorisation 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, and the other half were administered order 2 first.

3.3. Results and discussion

The results for the first and second aIAT administered were analysed separately and the dependent measure was the D-IAT (D600 algorithm; Greenwald et al., 2003). Also here, as before, we will report only analyses for this index as analysis conducted on raw RTs are comparable.

3.3.1. D-IAT

The D-IAT was submitted to a univariate ANOVA with group ('thieves' vs. 'readers') and order (order 1 vs. order 2) as between-subject factors. The D-IAT was 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 was expected for 'thieves' and a negative D-IAT was expected for 'readers'.

3.3.1.1. Analysis of the first aIAT. 'Thieves' showed, as expected, a positive D-IAT whereas 'readers' showed a negative D-IAT (0.68 vs. −0.43; $F(1,36) = 55.243$, $p<0.001$, $\eta^2 = 0.605$). Furthermore, order 1 showed a greater D-IAT effect than order 2 (0.30 vs. −0.05; $F(1,36) = 5.320$, $p = 0.024$, $\eta^2 = 0.133$).

3.3.1.2. Analysis of the second aIAT. 'Thieves' showed positive D-IAT whereas 'readers' showed negative D-IAT (0.58 vs. −0.31; $F(1,36) = 71.142$, $p<0.001$, $\eta^2 = 0.664$). As previously described, order 1 showed a greater D-IAT effect than order 2 (0.26 vs. 0; $F(1,36) = 5.675$, $p = 0.023$, $\eta^2 = 0.136$).

The correlation between the D-IAT calculated in the first test and the same index calculated in the second test was r = 0.63 ($p<0.001$). This index is not exactly test–retest reliability, given that the order of the congruent and incongruent blocks was reversed in the two aIAT administrations. This high correlation indicates, however, that the D-IAT is a stable measure, relatively independent of the order of presentation of the congruent block (in third position or in fifth position in the five blocks sequence).

To summarise the data collected on the mock-crime experiment, we originally found 93% accuracy in classifying guilty and innocent participants (Sartori et al., 2008). In a replication of the same experiment, Verschuere et al. (2008) found, however, a lower accuracy for the same experiment (64%). This inconsistency could be owed to the use of negative sentences and negative reminder labels. Here, we showed in Experiments A and B that the use of negative sentences and reminder labels reduces classification accuracy of individual participants compared with a similar test using only affirmative sentences and reminder labels. For this reason we 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 C, reported here, the participants were administered two aIATs, one with the congruent block as third block and the other with the congruent block as fifth block. Taken individually, both the first IAT and the second aIAT had a classification accuracy of 88%, averaging the two D-IATs we obtained accuracy in classifying individual participants of 95% for both innocent and guilty participants (Table 1).

4. General discussion

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

Here we report on five experiments to investigate systematically 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 90% in the experiments previously reported (Agosta et al., 2010). 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 from* instead of the negative (e.g. *I have been to Rome* vs. *I have been to a different place from Rome*). Negative and affirmative Counter-event sentences can be considered, from this point of view, as equivalent.

One possible explanation of the detrimental effect owing to the use of negative reminder labels and negative sentences refers to the figure-ground model of Rothermund and Wentura (2004). This model assumes that the IAT 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) rather than 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 reported by Brendl,

Markman, and Messner (2001) in which a strong association was found between insects and pleasant words while non-words were associated with unpleasant ones. This effect cannot be explained by implicit associations, Rothermund and Wentura (2004) say that the figure-ground model gives an explanation based on the salience of unpleasant words compared with pleasant and non-words compared with insects (Rothermund & Wentura, 2004). Similarly to the experiment reported by Brendl et al. (2001) in which negative concepts increase the saliency, here, negative and false sentences are assumed to have higher saliency. In our experiments, according to this view, faster RTs in the block pairing true sentences and event sentences (even when the event is false) may be owed to the greater salience of negative stimuli with respect to the affirmatives and by the greater salience of false stimuli with respect to the true ones (Rothermund & Wentura, 2004). Thus, the salience effect may be stronger than the association effect whenever negatives are used, leading to misdiagnosis when using the aIAT. As an example consider Experiment B1. Here, high false positive rates are the result of pairing TOKIO/TRUE and NOT TOKIO/FALSE and this is due to NON TOKIO being more salient than TOKIO and FALSE being more salient than TRUE.

In contrast, the mock-crime experiment (Experiment 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 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 the greatest care should be taken not to include two events that are both true and false. Furthermore, the two events should be described in an affirmative format and referred to by 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 reminder labels could be OPEN and CLOSED. Inadequate sentences are *I did not close the door* and an inefficient reminder label is NOT OPEN.

Developing an aIAT using Events (affirmative) and Counter-events (negative) to describe autobiographical memories orally is easy but leads, as reported here, to unreliable results.

Misdiagnosis, in this case, will magnify the classification of innocent subjects as guilty, exactly as reported by Verschuere et al. (2009), given that innocent subjects will have their memories referred to by the negative sentences (e.g. *I did not steal the CD*).

By contrast, the use of affirmative sentences for referencing contrasting autobiographical memories is more difficult to fine-tune if one keeps in mind that only one of the two memories must be true and the other false. In typical forensic applications, however, this will not be much of a problem, as two affirmative contrasting hypotheses are clearly stated (the prosecutor's hypothesis and the defence hypothesis), only one of the two being true.

Thus, here we showed that the use of negative sentences/labels does not give rise to reliable and replicable results, as the mock-crime experiment had different accuracy rates (93% in Sartori et al.'s paper vs. 64% in Verschuere et al.'s paper). In contrast, affirmative sentences give rise to reliable replicable highly accurate classifications (Card aIAT, Exp 1; Holiday aIAT, Exp 4; Sartori et al., 2008).

An interesting issue may arise when the aIAT is confronted with another lie-detection technique, the Guilty Knowledge Test (GKT; Lykken, 1960, 1998). The GKT uses multiple-choice questions, each including a 'relevant' answer (e.g. feature of the crime under investigation) and several 'control' answers that cannot be distinguished from the relevant answer by an innocent suspect (Lykken, 1998). Typically, guilty suspects exhibit greater physiological responses for relevant than for control alternatives. In the GKT, if an innocent suspect has been exposed to information about the crime it will result in an increase of physiological responses, resulting in a high rate of false positives. Thus, the innocent suspect will be classified as guilty. Ben-Shakhar and Elaad (2003) explain that avoiding leakage of critical items is crucial for a successful implementation of the GKT because the main advantage of the technique is the protection supplied to the innocent suspects, thus avoiding false-positives; the innocent suspect in fact can have access to critical information through newspapers or other media.

In this regard, the aIAT may not be affected by similar limitations when there is also a high risk of exposure to guilty information for innocent suspects. The aIAT may, at least in some conditions, be used as a valid substitute for the GKT, given its high accuracy and its similar misdiagnosis rate for guilty and innocent subjects. Experiment C shows that the aIAT is not affected by exposure to guilty information; in fact control subjects, who read an article regarding the crime and were therefore exposed to the same information as the guilty subjects who acted the mock crime, were correctly classified (95%) as innocents on the basis of the D-IAT.

In conclusion, the aIAT is an instrument that correctly identifies which one of two contrasting and mutually exclusive events is true for the subject given that negatives are avoided both when selecting reminders labels and sentences describing autobiographical events.

The cognitive mechanism underlying the aIAT is assumed to be similar to the one involved in the standard IAT effect (Greenwald et al., 1998) studied on concepts and, therefore, also here two dueling explanations may be put forward. The aIAT effect may arise because of strength of associations or may, by contrast, be the result of the differential saliency among the stimuli. The detrimental effect of negative sentences on aIAT diagnostic accuracy may not be easily accounted within the associative strength hypothesis while it can more easily be explained within the saliency hypothesis.

References

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (2010). Detecting fakers of the autobiographical IAT. *Accepted for Publication in Applied Cognitive Psychology*.

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.

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.

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197−216.

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General*, *134*, 420−425.

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

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

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

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

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

Verschuere, B., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking the autobiographical IAT. *Psychological Science*, *20*, 410−413.