# Application of the Implicit Association Test to a Study on Deception

PETER FROST, MICHAEL ADIE, ROLAND DENOMME, ANNABEL LAHAIE, and ANGELA SIBLEY
Southern New Hampshire University

EMILY SMITH
University of New Hampshire

Three experiments were conducted to find out whether the standard Implicit Association Test (IAT) could be used to distinguish truthful and deceitful witnesses. We anticipated that IAT effects would be greater after lying. Participants were asked to answer questions with incorrect answers (i.e., the lie condition) or correct answers (i.e., the truthful condition). A third group of participants were not interviewed (a control group). Participants then took the IAT, in which they were asked to associate correct and incorrect answers with positive or negative attributes. Results demonstrate that standard IAT effects are greater after lying than after truth telling, but only when attribute labels were clearly and explicitly linked to positive and negative affect. Theoretical implications are considered.

Psychologists have long been interested in whether people behave in different ways when they are telling a lie and when they are telling the truth. Psychological research on cues associated with deception has been conducted over several decades. However, research on cues to deception has increased in recent years, given intense interest in issues of national security and terrorism (see DePaulo et al., 2003, for a review of the literature on cues to deception). In this article we define *deception* as the act of deliberately providing false information. Although some researchers draw a distinction between *deception* and *lies* (e.g., Bok, 1978), we use these words interchangeably.

Although studies on cues to deception have long focused on links to deception associated with arousal and feelings (DePaulo et al., 2003; Zuckerman, DePaulo, & Rosenthal, 1981), many recent studies have explored the cognitive aspects of deception (Gregg, 2007; Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008; Walczyk, Roper, Seemann, & Humphrey, 2003). For example, Walczyk et al. (2003) claimed that the cognitive process involved in telling lies is more complex than that associated with telling truths. Reaction times (RTs) associated with open-ended questions tended to be longer when telling a lie than when telling a truth. Walczyk et al. speculated that the longer times associated with lying occurred because of the additional cognitive processes associated with

American Journal of Psychology
Summer 2010, Vol. 123, No. 2 pp. 221–230 • © 2010 by the Board of Trustees of the University of Illinois

the decision to lie and the construction of a lie (see also Spence et al., 2001).

Gregg (2007) used another reaction time–based lie detection test called the Timed Antagonistic Response Alethiometer (TARA). The TARA requires the speeded classification of sentences into different categories—such as true and false—as accurately as possible. Participants who have lied are required to perform two incompatible classification tasks, whereas participants who tell the truth perform two compatible classification tasks. As a consequence, dishonest respondents were forced to perform more slowly than honest respondents in order to achieve equivalent levels of accuracy. Gregg found that the TARA could distinguish groups of liars and truth tellers with 85% accuracy.

We were interested in whether the Implicit Association Test (IAT), another cognitive RT paradigm, could be used as a tool to distinguish deception from truth telling. The IAT is a task, originally developed by Greenwald, McGhee, and Schwartz (1998), in which speed of response is intended to assess the relative strength with which attitudes are associated with positive and negative evaluations. Participants classify stimuli representing two target concepts (e.g., a lie vs. a truth) and identify attributes as good or bad words (e.g., *happiness* vs. *death*) by pressing buttons on computer keyboard.

The IAT has traditionally been used to assess implicit racial attitudes. For example, participants might see a series of pictures of either African Americans or Caucasians, followed by words that are either positive or negative. Participants must indicate whether the word is positive or negative as quickly as possible. Participants often respond more quickly and accurately if good words share the same response key with pictures of Caucasian faces and bad words share the same key with pictures of African American faces than vice versa (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Greenwald et al., 1998).

The IAT is based on the assumption that if the object of an attitude elicits a particular evaluation, whether positive or negative, it will affect RTs to other congruent and co-occurring stimuli. RT is used to determine the strength of association between some object of an attitude and evaluation of that object (Dasgupta & Greenwald, 2001). Several published studies confirm the construct and predictive validity of the IAT as legitimate measure of implicit attitudes (see Dasgupta & Greenwald, 2001, for a summary of studies on the validity of the IAT).

Although the IAT was intended to measure attitudes or feelings associated with certain people, events, or things, some studies have found evidence that the IAT might measure other variables. For example, Rothermund and Wentura (2004) investigated whether IAT effects are influenced by salience asymmetries rather than associations. They were able to demonstrate that IAT effects were dependent on salience asymmetries while holding associations between categories constant. Although Rothermund and Wentura offered a different account of the IAT from that originally posited, they acknowledged that the IAT, as an experimental paradigm, "reveals strong, replicable, and nontrivial interference effects" (p. 159).

Although the mechanism behind implicit association effects has yet to be fully resolved, some interesting findings have recently appeared in the application of the IAT to lie detection. A new method, the autobiographical Implicit Association Test (aIAT), is used to assess the truthfulness of autobiographical statements (Sartori et al., 2008). Participants classify sentences seen on a computer screen by pressing one of two response keys as quickly as they can in one of two tasks. In the confession–true task, confessions (e.g., "I have stolen the watch") and true statements are associated with one key, and denial (e.g., "I did not steal the watch") and false statements are associated with the other key. In the denial–true task, confessions and false statements are associated with one key, and denial and true statements are associated with the other key. Sartori et al. found that guilty participants were faster in the confession–true task than in the denial–true task, whereas innocent participants showed the opposite results. Sartori found that the aIAT could distinguish groups of liars and truth tellers with 91% accuracy (higher than that associated with TARA).

In this study, we wanted to find out whether the standard IAT could be used to find out whether participants were engaging in deception. Participants were asked to answer questions during an interview with incorrect answers (the deception condition) or correct answers (the truthful condition). Participants then responded to the IAT. In one block they were asked to press one key for correct answers or words associated with positive affect and a different

key for incorrect answers or words associated with negative affect. In another block, participants were asked to press one key for correct answers or words associated with negative affect and a different key for incorrect answers or words associated with positive affect. We hypothesized that participants in the deception condition would be slower at following rules that paired correct answers with negative affect and incorrect answers with positive affect than the reverse rule. Theoretically, the association and salience accounts would both be consistent with this predicted outcome. If the association account is true, then participants in the deceptive condition should be slower than participants in the truth condition at associating incorrect answers with positive affect because lying might elicit more of a negative reaction to this pairing. If the salience account is true, then participants in the deception condition should experience more of an interference effect because lying might facilitate perception of a salience mismatch.

The goal of this study was to find out whether the standard IAT (sIAT) would reveal a greater effect after lying than after truth telling. If the hypothesis is confirmed, the sIAT could become a new paradigm in exploring the cognitive aspects of lying. The results of this study could have implications for the theoretical mechanisms of the IAT.

## EXPERIMENT 1A

Participants in Experiment 1a responded to the IAT, in which they were asked to associate deceptive or truthful items from the interview with positive or negative attributes.

### METHOD

#### *Participants*

Participants were 51 undergraduate students (41 women and 10 men) from Southern New Hampshire University who were recruited on a voluntary basis. Participants were from upper- and lower-level psychology classes. Students were offered extra course credit for their participation.

#### *Materials*

##### *INTERVIEW QUESTIONS*

Participants were handed a list of eight questions, each followed by a correct and an incorrect answer (see Appendix A). All questions were about Southern New Hampshire University.

##### *IAT*

The IAT is intended to be a test of automatic attitudes. The IAT was applied to this study to see whether participants harbor a negative implicit association with items they were asked to use during a deceptive task. The eight incorrect and eight correct answers were shown during the IAT. The evaluative attributes were eight positive (*marvelous, superb, pleasure, beautiful, joyful, glorious, lovely,* and *wonderful*) and eight negative (*tragic, horrible, agony, painful, terrible, awful, humiliate,* and *nasty*) attributes.

The IAT was administered on Dell XPS 410 desktop computers with 20-in. monitors using Windows XP operating systems. The software used to administer the IAT was Inquisit 2.0 by Millisecond.

#### *Procedure*

##### *STUDY AND INTERVIEW PHASE*

No more than one participant was administered the experiment at a time. Participants entered a room and sat in a chair facing the experimenter. Participants first engaged in a study phase and then an interview phase, each conducted by a different experimenter. During the study phase, participants were handed a list of eight questions, each with a correct and an incorrect answer. They were told to study both the correct and incorrect answers. Participants had 4 min to study all potential answers but were allowed more time if they requested it, with no more than 8 min of total study time. Before the study phase started, participants were randomly placed in one of three conditions (between participants, $N = 17$ for each condition): no interview, truth, and deception conditions. Participants in the no-interview condition studied the correct and incorrect answers but were not interviewed (a control condition). There were two interview conditions: truth and deception. Participants in the truth condition studied the correct and incorrect answers during the study phase and then answered the eight questions with the correct answers orally during a brief interview (a second control condition). Participants in the deception condition studied the correct and incorrect answers during the study phase and then answered the eight questions with the incorrect answers orally during a brief interview (the treatment condition). The first experimenter in the truth condition ended the study phase by instructing the participants to tell the truth during the interview with the next experimenter. The first experimenter in the deception condition ended the study phase by instructing participants to provide incorrect answers during the interview with the next experimenter, even though the second interviewer

told them to provide the correct item. Instructions regarding the procedure of the experiment found in the informed consent also reinforced the idea that participants should answer questions during the interview with the answers specified by the first experimenter (if they should find themselves in one of the interview conditions). During the interview, the participant and interviewer sat in chairs facing one another with about 1.22 m between chairs. Interviewers read each question from the study phase and made eye contact with the participants when they answered the questions.

*IAT PHASE*

After the interview (or after the study phase in the case of participants in the no-interview condition), participants responded to the IAT on the computer. Participants viewed the correct and incorrect answers from the study phase. In one phase, participants associated incorrect items with negative attributes using a certain response key (e.g., the "E" key) and correct items with positive attributes using a different response key (e.g., the "I" key). In another phase participants did the opposite; they associated incorrect items with positive attributes using a certain response key and correct items with negative attributes using a different response key. The phases were counterbalanced so that the order of these phases was used equally often with an equal number of participants in each sequence of phases. The IAT consisted of five separate blocks. See Table 1 for a description of the blocks and counterbalancing schemes. Blocks 1, 2, and 4 included 32 trials. Blocks 3 and 5 included 96 trials. RTs were measured in milliseconds. A correct or incorrect answer from the study phase was shown at the center of the screen, and the rules were shown at the top of the screen.

*Design*

The experimental design used was a 3 (type of interview: truth, lie, or no interview) × 2 (IAT combinations: correct information + pleasant vs. incorrect information + pleasant) factorial design. Type of interview was varied between participants, and IAT combinations were varied within participants.

## RESULTS AND DISCUSSION

The answers recorded during the two interview conditions were analyzed in order to assess the extent to which participants recalled studied items and followed directions. Participants in the truth condition responded with correct information 93% of the time on average ($SD$ = .72). Participants in the deception condition responded with incorrect information 90% of the time on average ($SD$ = .76).

IAT data were analyzed using the scoring algorithm, hereafter referred to as $D$, described by Greenwald, Nosek, and Banaji (2003). The $D$ index expresses the IAT effect—the difference in RTs between the two double-categorization blocks (Blocks 3 and 5)—relative to the standard deviation of the latency measures. More specifically, $D$ is calculated by subtracting the corrected mean RT (see Greenwald et al., 2003) for Block 3, associating correct and good–incorrect and bad, from the mean RT for Block 5, associating correct and bad–incorrect and good, and then dividing this difference by the inclusive standard deviation of the two blocks.

Before calculating $D$, we discarded RTs shorter than 300 ms or longer than 10,000 ms. Latencies were log-transformed in order to normalize the distribution and meet the assumptions for the inferential statistics used in this study.

IAT effects associated with the two answer combinations (correct information and pleasant–incorrect information and unpleasant vs. incorrect information and pleasant–correct information and unpleasant), broken down by interview type, are shown in Figure

**TABLE 1.** Blocks used for the standard Implicit Association Test

| Response key | Block 1: Accuracy discrimination | Block 2: Attribute discrimination | Block 3: Accuracy + attribute combined | Block 4: Reversed accuracy discrimination | Block 5: Reversed accuracy + attribute combined |
|---|---|---|---|---|---|
| "E" | Correct | Good | Correct and good | Bad | Correct and bad |
| "I" | Incorrect | Bad | Incorrect and bad | Good | Incorrect and good |

Note. *The orders of Blocks 3 and 5 and of Blocks 2 and 4 were reversed for half the participants.*

1. The IAT effects associated with the interview types of deception, truth, and no interview were 399 ms ($D$ = .99), 425 ms ($D$ = 1.12), and 378 ms ($D$ = .91), respectively. A main effect was observed for IAT combinations, $F(1, 48) = 213.06$, $p_{rep} = .96$, $\eta^2 = .08$. Interview type and a combination × interview type interaction were not significant, $F$s < 1.

Results indicated that participants were faster when rules required them to pair correct information with "good" words and incorrect information with "bad" words than the other way around. However, the magnitude of these IAT effects was similar in all three conditions (the lie, truth, and no-interview conditions).

## EXPERIMENT 1B

In addition to replicating Experiment 1a, Experiment 1b included a measure of explicit (i.e., self-reported) attitudes. Past studies have demonstrated that implicit and explicit measures of attitude sometimes differ because participants have more time to reflect and adjust their answers to explicit reactions (Dasgupta & Greenwald, 2001). In addition, we avoided the bias that might have been inherent in using the terms *incorrect* and *correct* by referring instead to "Response A" and "Response B" items during all phases of the experiment. For example, participants in Experiment 1a might have been slower to respond to incorrect items because they were aware that the term *incorrect* has a negative connotation. By referring to the more objective terms *Response A* and *Response B* on the computer screen, and having each represent correct items half the time and incorrect items half the time, we intended to avoid any negative reactions associated with the attribute labels.

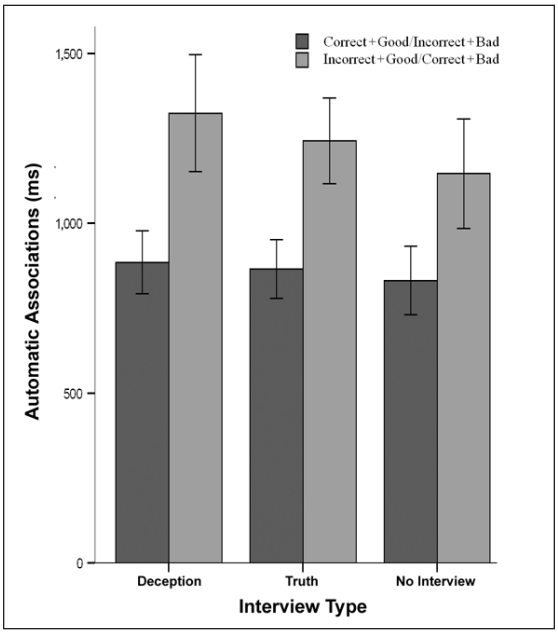


**FIGURE 1.** Effects of interview type on Implicit Association Test effects, Experiment 1a

### METHOD

#### *Participants*

Fifty-one undergraduate students (39 women and 12 men) from Southern New Hampshire University volunteered for course credit. Participants were from lower- and upper-level psychology classes.

#### *Materials, Procedure, and Design*

With a few exceptions, the number of participants in each condition ($N$ = 17), materials, procedure, and design were the same as those in Experiment 1a. Experiment 1b included the following exceptions: Correct and incorrect answers were referred to as Response A and Response B, respectively, or vice versa. Response Set A represented incorrect items for half of all participants and correct for the remaining participants. The same was true for Response Set B. Response A and Response B were shown in place of correct and incorrect answers within the IAT. We included "feeling thermometers" as ratings of explicit attitude toward the answers studied during the interview (see Dasgupta & Greenwald, 2001, for more details). Participants placed an $X$ on a picture of a thermometer at a point that expressed their attitude toward a particular word from the interview, from 0° (*cold or unfavorable*) to 100° (*warm or favorable;* 50° = *neutral*). Another procedural difference was that we asked participants to memorize which answers were associated with Response A and Response B during the study phase. Participants in the truth condition were asked by the first experimenter to respond with the same response set as that requested during the interview, whereas participants in the deception condition were asked by the first experimenter to respond with the response set that was not requested during the interview.

### RESULTS AND DISCUSSION

Answers recorded during the interview were analyzed. Participants in the truth condition answered questions with response set items requested during

the interview 92% of the time on average ($SD$ = .68). Participants in the deception condition responded with response set items not requested during the interview 94% of the time on average ($SD$ = .72).

IAT data were prepared and transformed in the same way as in Experiment 1a (Figure 2). IAT latencies were analyzed using the algorithm by Greenwald et al. (2003). IAT effects for the deception, truth, and no-interview conditions were 475 ms ($D$ = 1.08), 365 ms ($D$ = 1.01), and 290 ms ($D$ = .95), respectively. A significant main effect was observed for IAT combinations, $F(1, 48) = 532.06$, $p_{rep} = .93$, $\eta^2 = .09$. Neither a main effect for interview type nor a combination by interview type interaction was observed, $F$s < 1.

Explicit measures of attitudes toward answers during the interview were measured using feeling thermometers. The higher the ratings, the more favorable was the participant's attitude toward the answer. When we analyzed incorrect answers only, a difference in average ratings was found for type of interview: truth ($M$ = 44.20°), lie ($M$ = 32.85°), and no interview ($M$ = 49.25°), $F(1, 48) = 4.26$, $p_{rep} = .82$, $\eta^2 = .18$. A Tukey HSD test indicated that the difference was between the lie condition and the truth and no-interview conditions, $p_{rep} = .74$. An analysis of the correct answers revealed no difference between truth ($M$ = 57.03°), lie ($M$ = 69.59°), and no-interview ($M$ = 65.16°) conditions, $F$ < 1. Participants rated correct answers more favorably ($M$ = 64.69°) than incorrect answers ($M$ = 44.17°), $t(50) = 6.82$, $p_{rep} = .86$, $d = .62$.

The results of Experiment 1b replicated the finding that participants are slower responding to rules that pair incorrect information with "good" attributes than with "bad" attributes. The magnitude of these IAT effects toward incorrect information was not found to differ significantly between the three types of interview conditions (the lie, truth, and no-interview conditions). Explicit ratings of attitudes toward incorrect answers, on the other hand, were affected by the type of interview, with ratings indicating that the least favorable attitudes were for answers associated with the lie condition.

## EXPERIMENT 2

Positive and negative affect associated with correct and incorrect information is only one possible explanation for the results of Experiments 1a and 1b.

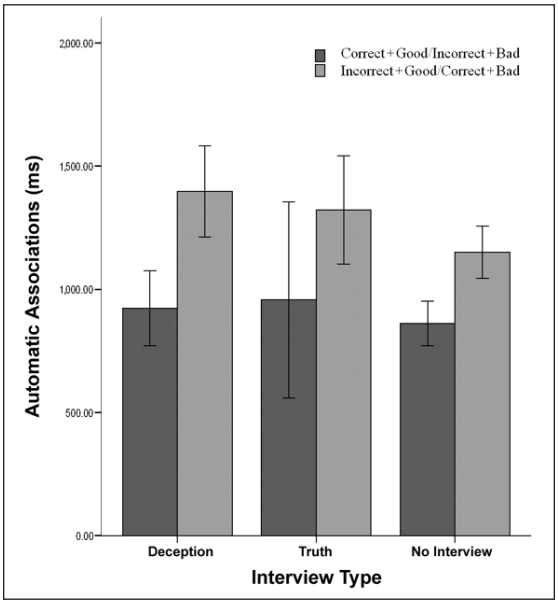


**FIGURE 2.** Effects of interview type on Implicit Association Test effects, Experiment 1b

The attribute labels *good* and *bad* might have been ambiguous in these experiments. For example, correct information might have been perceived as good information and incorrect information as bad information. As a consequence, *good* and *bad* attribute labels might not have been associated with negative or positive affect. Rather, they would be linked with the accuracy of information, an explanation hereafter referred to as the accuracy account. In Experiment 2, we manipulated the attribute labels used. In one condition, we used *good* and *bad,* as before; in the other condition, we used *positive affect* and *negative affect*.

We also set up Experiment 2 so that stimuli were correct for some participants but incorrect for others (using a crime scene description with two different sets of target items). In this way, we were able to determine whether the standard IAT is useful in assessing the accuracy of eyewitness testimonies and distinguishing liars from truth tellers.

### METHOD

#### *Participants*

The experiment included 108 undergraduate students (80 women and 28 men) from Southern New Hampshire University who were recruited on a voluntary

basis. Participants were offered extra course credit for upper- and lower-level psychology classes.

### Materials

#### CRIME SCENE DESCRIPTION

Participants were handed paragraphs describing a modified version of an actual police confrontation that took place in New York City (see Appendix B), based on a *New York Magazine* article (Rovzar, 2009). Six items in the description were critical items (italicized in Appendix B). Notice that in Appendix B, there are two versions of each of the six critical items. One of the versions was associated with Response A, and the other was associated with Response B. Participants were given one sheet with "Response A" at the top and another with "Response B" at the top. Each sheet showed the same crime scene description but with only one version of the critical items. Each version of the critical items served equally often as Response A and Response B items.

#### IAT

The IAT was administered using the same equipment and software as in Experiments 1a and 1b. Response A items and Response B items were shown during the IAT. The evaluative attributes were six positive and six negative attributes randomly sampled from Experiment 1a. Half the participants were exposed to the "Good" and "Bad" attribute labels, whereas the other half were exposed to the "Positive Affect" and "Negative Affect" attribute labels.

### Procedure

#### STUDY AND INTERVIEW PHASE

Participants were asked to study the crime scene with Response A critical items and then Response B critical items, or vice versa (order of presentation of each response set was counterbalanced). Participants were told that one of the response sets was correct after the study phase was complete. Half of the participants were randomly assigned to the Response A correct condition, and the remaining participants were assigned to the Response B correct condition. Participants were told to keep track of which items belonged to Response Set A and Response Set B. Participants had 5 min to study each response set. Participants were placed in one of three conditions (between participants, $N = 18$ for each condition in both the "Good" vs. "Bad" and "Positive Affect" vs. "Negative Affect" conditions): no-interview, truth, and deception conditions. Participants in the no-interview condition were not interviewed (a control condition). There were two interview conditions, conducted in the same way as in Experiments 1a and 1b: the truth and deception conditions. Again, one experimenter provided the instructions for the study phase, and a different experimenter conducted the interview for the truth and deception conditions. The first experimenter in the truth condition instructed the participant to respond with the response set asked for by the interviewer. The first experimenter in the deception condition instructed the participant to respond with the response set that was not asked for by the interviewer (e.g., the first experimenter might instruct the participant to respond with Response A items, whereas the interviewer asks for Response B items).

#### IAT PHASE

The IAT was run in the same way as in Experiment 1b, except that half of the participants were randomly exposed to the "Good" and "Bad" attribute labels, and the other participants saw the "Positive Affect" and "Negative Affect" attribute labels.

#### DESIGN

The experimental design used was a 3 (type of interview: truth, lie, or no interview) × 2 (type of attribute label: "Good" and "Bad" or "Positive Affect" and "Negative Affect") × 2 (response type: Response Set A correct vs. Response Set B correct) factorial design. Type of interview, type of attribute label, and response set declared correct were varied between participants.

## RESULTS AND DISCUSSION

As in Experiments 1a and 1b, answers recorded during the two interview conditions were analyzed. We first analyzed data in the "Good" and "Bad" attribute condition. Participants in the truth condition responded with the response set requested by the interviewer (i.e., responding with Response Set A items in the Response Set A correct condition or responding with Response Set B items in the Response Set B correct condition) 90% of the time on average ($SD = .76$). Participants in the deception condition responded with the response set opposite of that requested by the interviewer (responding with Response Set B items in the Response Set A correct condition or responding with Response Set A items in the Response Set B correct condition)

88% of the time on average (*SD* = .80). We then analyzed interview answers in the "Positive Affect" and "Negative Affect" attribute condition. Participants in the truth condition responded with response set items consistent with that requested by the interviewer 92% of the time on average (*SD* = .68). Participants in the deception condition responded with response set items inconsistent with that requested by the interviewer (the response set opposite of that requested by the interviewer) 89% of the time on average (*SD* = .74).

We then analyzed the IAT data. RTs shorter than 300 ms or longer than 10,000 ms were discarded. Data are shown in Table 2. Positive *D* scores indicate a faster response when Response A was paired with the label "Positive Affect" or "Good" than "Negative Affect" or "Bad." Negative *D* scores indicate a faster response when Response B was paired with the label "Positive Affect" or "Good" than "Negative Affect" or "Bad." A hit occurred when a participant had a positive *D* score in the Response A correct condition or a negative *D* score in the Response B correct condition.

Across all conditions in Experiment 2, IAT effects were more positive for the Response A correct condition than the Response B correct condition, $F(1, 102) = 311.21$, $p_{rep} = .93$, $\eta^2 = .07$. A main effect for interview type and an interview type × response type interaction did not occur, $F$s < 1. However, there was a significant three-way interaction of interview type × response type × attribute label, $F(2, 102) = 3.02$, $p_{rep} = .97$, $\eta^2 = .05$. The two-way interaction of interview type and response type was significant for the "positive affect" versus "negative affect" label condition but not for the "Good" versus "Bad" label condition, as revealed by Tukey HSD tests.

These results demonstrate that standard IAT effects are greater after lying than after truth telling, but only when attribute labels are unambiguously linked to positive and negative affect. The mere act of lying appears to slow down reaction to any rule that requires pairing correct information and positive affect or incorrect information and negative affect.

## GENERAL DISCUSSION

Taken together, our results indicated that IAT effects can distinguish correct from incorrect answers. Moreover, if the attribute labels are clearly associated with positive and negative affect, IAT effects are even greater after lying than after truth telling (Experiment 2). These findings are consistent with other studies, which have demonstrated that a similar version of the IAT, namely the aIAT, can distinguish deceptive witnesses from truthful ones (Sartori et al., 2008; Verschuere, Prati, & De Houwer, 2009). This series of experiments extends these findings to reveal that the sIAT for correct and incorrect information can also distinguish deceptive and truthful witnesses.

We originally suspected that Experiments 1a and

**TABLE 2.** Mean Implicit Association Test effects (*D*s), reaction times (RTs, in ms), standard deviations, and hit rates by type of interview

| | Type of interview | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Deception | | | Truth | | | No interview | | |
| | *D* | RT (*SD*) | Hit rate (%) | *D* | RT (*SD*) | Hit rate (%) | *D* | RT (*SD*) | Hit rate (%) |
| | "Positive Affect" versus "Negative Affect" | | | | | | | | |
| Response A correct | +.81 | 517 (.56) | 95 | +.52 | 372 (.54) | 67 | +.50 | 364 (.47) | 63 |
| Response B correct | –.88 | 524 (.51) | 92 | –.49 | 369 (.48) | 65 | –.47 | 362 (.46) | 64 |
| | "Good" versus "Bad" | | | | | | | | |
| Response A correct | +.54 | 378 (.61) | 70 | +.49 | 359 (.58) | 67 | +.45 | 352 (.50) | 65 |
| Response B correct | –.50 | 370 (.55) | 74 | –.52 | 368 (.46) | 73 | –.43 | 349 (.52) | 66 |

Note. *Positive* D *scores indicate a faster response when Response A was paired with the label "Positive Affect" or "Good" than "Negative Affect" or "Bad." Negative* D *scores indicate a faster response when Response B was paired with the label "Positive Affect" or "Good" than "Negative Affect" or "Bad." A hit occurred when a participant had a positive D score in the Response A correct condition or a negative D score in the Response B correct condition.*

1b demonstrated that participants held more negative attitudes (implicitly) toward incorrect information than correct information. The results of Experiment 2 revealed that the accuracy account—the association of the attribute labels "Good" and "Bad" with good and bad information rather than good and bad affect—might explain the results of the first two experiments. Once the attribute labels were more explicitly associated with feelings rather than accuracy, lying led to larger IAT effects than telling the truth (or after not being interviewed at all). Although this finding is consistent with the affective account, we cannot rule out some other potential explanations. For example, the results might also reflect greater salience of incorrect information after lying and a salience mismatch across blocks of the IAT (Rothermund & Wentura, 2004). Nonetheless, the possibility of strong interference effects caused by differences in salience of truthful and deceptive information would also be worthy of further investigation.

Even if an affective account is supported by future research, further understanding of why a negative implicit attitude toward deceptive information might increase RTs must be explored. As stated earlier, some researchers, such as Walczyk et al. (2003), have found that RTs when lying tend to be longer than when telling the truth. They speculated that this occurs because liars must actively inhibit the inclination to tell the truth. Spence at al. (2001) found that when participants answered questions deceptively, the regions of the brain associated with response inhibition (the bilateral and ventrolateral prefrontal cortices of the brain) were more active than when they told the truth. It is conceivable that implicit associations for lies, and inaccurate information in general, are slower than for truths because participants additionally process, and perhaps inhibit, thoughts associated with the consequences of false information.

The explicit attitude ratings of Experiment 1b indicated a more negative attitude toward incorrect information than correct information. Furthermore, ratings for answers that were previously associated with telling lies were higher than for answers previously associated with telling the truth. This finding is consistent with past research that has revealed at least some dissociation between tests of implicit and explicit attitudes (Dasgupta & Greenwald, 2001; Greenwald & Banaji, 1995).

Although the construct and predictive validity of the IAT are well established for its use in detecting implicit racial attitudes in most studies (Dasgupta et al., 2000; Greenwald et al., 1998; McConnell & Leibold, 2001, but see Kinoshita & Peek-O'Leary, 2005), future research will need to determine the validity of applying the IAT to the detection of deceptive and false answers. Test administrators would need to be particularly careful to prevent malingering when applying the IAT to forensic situations. Kim (2003) found that when participants are instructed to fake their attitudes on the IAT they can do so effectively, but only if they are given specific instructions on how to do so. Verschuere, Prati, and Houwer (2009) showed that, like other lie detection tests, the aIAT is vulnerable to faking. They found that participants guilty of a mock theft were able to obtain innocent test outcomes if they were instructed—and taught how—to fake. Faking might also affect use of the standard IAT for lie detection purposes.

The sIAT can be a strong indicator of incorrect eyewitness accounts, particularly after deception has occurred. These findings suggest merely one of potentially many ways the IAT can be used when detection of malingering is of concern.

*APPENDIX A. QUESTIONS*

1. What is the name of the library at Southern New Hampshire University?
   **Incorrect Answer:** Regina Library
   **Correct Answer:** Shapiro Library
2. What appears in the Southern New Hampshire University athletic logo?
   **Incorrect Answer:** Bear
   **Correct Answer:** Penman
3. What is the color of the leather chairs out in the hallways and lobbies of Robert Frost Hall?
   **Incorrect Answer:** Red
   **Correct Answer:** Tan
4. The observation room for psychology (in Robert Frost Hall) is located where?
   **Incorrect Answer:** First floor, Walker Auditorium
   **Correct Answer:** Third floor, RF315
5. In what city is Southern New Hampshire University located?
   **Incorrect Answer:** Houston
   **Correct Answer:** Manchester
6. In which month does school begin at Southern New Hampshire University?
   **Incorrect Answer:** August
   **Correct Answer:** September

7. What are the school colors of Southern New Hampshire University?
   **Incorrect Answer:** Green and gold
   **Correct Answer:** Blue and yellow
8. What brand of computers is used in the first floor lab in Robert Frost Hall?
   **Incorrect Answer:** Macintosh
   **Correct Answer:** Dell

### APPENDIX B. CRIME SCENE DESCRIPTION, EXPERIMENT 2

Yesterday a man was handcuffed [in *Times Square/outside of Yankee Stadium*] for wandering around in a [*Superman/Batman*] costume. Police approached [*Maksim "Clark Kent" Katsnelso/Frank "Bruce Wayne" Frisoli*] and asked him if he had "the required license to perform in costume in public." When he said he didn't, the police tried to [*taser/handcuff*] him. A witness claimed that it took [*2 officers/8 officers*] to hold down the "super hero." He was putting up a good fight. A child at the scene said, "Mommy, it's [*Superman/Batman*]!" Eventually, he allegedly [*punched a female cop/punched a male cop*] in the face and was taken to the station for assault and resisting arrest.

### NOTES

This study was funded by two research grants provided by Southern New Hampshire University. Funding for the software was provided by Southern New Hampshire University's School of Liberal Arts. We would like to thank two lab assistants, Kristin Culver and Stacy Rivard, for their help running the experiments. We also wish to thank Drs. Bruno Verschuere and Robert Proctor for their helpful comments on the earlier versions of the manuscript and direction of the second experiment.

Address correspondence about this article to Peter Frost, Department of Psychology, Southern New Hampshire University, 2500 North River Road, Manchester, NH 03106-1045 (e-mail: p.frost@snhu.edu).

### REFERENCES

Bok, S. (1978). *Lying: Moral choice in public and private life.* New York: Pantheon.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800–814. doi:10.1037/0022-3514.81.5.800

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Ruling out the familiarity effect. *Journal of Experimental Social Psychology, 36,* 316–328.

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

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: 1. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Gregg, A. P. (2007). When vying reveals lying: The Timed Antagonistic Response Alethiometer. *Applied Cognitive Psychology, 21*(5), 621–647. doi:10.1002/acp.1298

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

Kinoshita, S., & Peek-O'Leary, M. (2005). Does the compatibility effect in the race Implicit Association Test reflect familiarity or affect? *Psychonomic Bulletin & Review, 12*(3), 442–452.

McConnell, A. R., & Leibold, J. M. (2001). Relations between the Implicit Association Test, explicit racial attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology, 37,* 435–442.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*(2), 139–165. doi:10.1037/0096-3445.133.2.139

Rovzar, C. (2009, July 10). Apparently it is a crime to wear a costume in Times Square. *New York Magazine*. Retrieved from http://nymag.com/daily/intel/2009/07/apparently_it_is_a_crime_to_we.html

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

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioral and functional anatomical correlates of deception in humans. *NeuroReport, 12,* 2849–2853.

Verschuere, B., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking in the autobiographical IAT. *Psychological Science, 20*(4), 410–413.

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology, 17,* 755–774.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). New York: Academic Press.