Röhner, Jessica; Schröder-Abé, Michela; Schütz, Astrid

What do fakers actually do to fake the IAT? : An investigation of faking strategies under different faking conditions

Date of secondary publication: 07.04.2025

Accepted Manuscript (Postprint), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1073799

Primary publication

Röhner, Jessica; Schröder-Abé, Michela; Schütz, Astrid (2013): What do fakers actually do to fake the IAT? : An investigation of faking strategies under different faking conditions, in: Journal of research in personality, Amsterdam [u.a.]: Elsevier, Vol. 47, Nr. 4, pp. 330–338, doi: 10.1016/j.jrp.2013.02.009.

What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions

Jessica Röhner[a,*,1], Michela Schröder-Abé[b,1], Astrid Schütz[b]

[a] Department of Psychology, Chemnitz University of Technology, D-09107 Chemnitz, Germany
[b] Department of Psychology, University of Bamberg, Germany

A B S T R A C T

We studied strategies (acceleration, slowing down, committing errors, avoiding errors) that participants use to fake the IAT under different conditions (faking high vs. faking low scores; naïve vs. informed faking) and investigated which strategies lead to faking success. Naïve participants successfully faked low scores by slowing down on the congruent block and faked high scores by accelerating on that block. They (unsuccessfully) tried to fake by increasing errors. When participants had been informed about strategies, they slowed down on the incongruent block to fake high scores and slowed down on the congruent block to fake low scores. The results are discussed with respect to recently published indices to detect and correct faked IAT scores and highlight problems with such procedures.

1. Introduction

Recent research has shown that the IAT can be faked under certain conditions (e.g., De Houwer, Beckers, & Moors, 2007; Fiedler & Bluemke, 2005; McDaniel, Beier, Perkins, Goggin, & Frankel, 2009; Röhner, Schröder-Abé,& Schütz, 2011; Steffens, 2004). Nevertheless, little is known about *how* people fake the IAT. What strategies do they use and does the choice of strategy depend on whether they want to fake low scores or high scores? Do all participants use the same strategies regardless of any pre-existing knowledge they may have about how to fake the IAT? Given that IATs are now widely used in basic and applied research, understanding the processes behind faking has increased in importance. The present research aimed to investigate the strategies that fakers use depending on faking direction and previous knowledge. We also investigated whether those strategies are connected to faking success or represent only faking attempts.

1.1. IAT procedure

The IAT is a computerized sorting task developed by Greenwald, McGhee, and Schwartz (1998). Its goal is to assess the strength of implicit associations between two target concepts and an attribute dimension by measuring participants' reaction times. For the IAT, respondents must sort stimuli that appear consecutively in the middle of the computer screen into four different categories: (a) two contrasted target concept categories that form the target dimension and (b) two contrasted attribute categories that form the attribute dimension. In the extraversion IAT used in this study, the target dimension includes self-relevant vs. non-self-relevant words (e.g., me vs. other), and the attribute dimension includes extraversion-related vs. introversion-related words (e.g., talkative vs. shy). The IAT consists of seven blocks altogether, of which Blocks 1, 2, and 5 are the so-called single or practice blocks that introduce the target or attribute dimensions. In these blocks, the categories of either the target concepts or the attribute concepts are presented in the upper corners of each side (left and right) of the display screen. Participants are instructed to respond to exemplars of each category (appearing in the middle of the screen) by pressing a key that corresponds to the same side as the appropriate label (e.g., the letter *d* might be used to correspond to the word in the upper left corner and the letter *k* might be used to correspond to the word in the upper right corner). Blocks 3 and 4 as well as 6 and 7 are the so-called combined blocks in which the discrimination of the attribute is paired with the discrimination of the target (i.e., participants must assign words from all four categories in these blocks). In Blocks 3 and 4, extraversion-related and self-relevant words appear on the same side (and thus share a key) and introversion-related and non-self-relevant words appear on the other (and share the other key). In Blocks 6 and 7, participants must respond to introversion-related and self-relevant words with one key and to extraversion-related and non-self-relevant words with the other key.

The rationale behind the IAT is that the sorting task should be easier and thus completed more quickly when the two concepts

* Corresponding author. Fax: +49 (0)371 531 27459.
E-mail address: jessica.roehner@psychologie.tu-chemnitz.de (J. Röhner).
[1] These authors contributed *about equally* to this work and therefore share first authorship.

that share one response key are strongly associated. If two concepts are only weakly associated, sorting them into one category should be more difficult and therefore be conducted more slowly. The IAT effect is computed as the difference in reaction times between the two combined blocks (in the present study, self linked with extraversion vs. self linked with introversion) divided by their overall standard deviation. The IAT effect is used as an indicator of the strength of the association between the concepts (in our case, self and extraversion).

1.2. Which faking strategies are conceptually possible?

When considering the strategies that participants might use to fake the IAT in the desired direction, we took a deductive approach and analyzed which strategies could affect IAT scores. We considered conceptually possible strategies that involved the manipulation of reaction times and errors.

Because the IAT measures reaction times in categorizing sets of stimuli, the scores can be affected by specifically slowing down or speeding up on the combined blocks. Thus, when the goal is to present categories as strongly associated and they share the same key, one can react more quickly to fake scores in this direction, and when these categories require pressing different keys, one can react more slowly.

Imagine two fictitious persons: Hugh and Lisa. *H*ugh wants to fake *high* scores on extraversion (i.e., appear extraverted) and *L*isa wants to fake *low* scores (i.e., appear introverted). To fake *high* scores, Hugh could respond more quickly when self-relevant and extraversion-related words (and when non-self-relevant and introversion words) require pressing the same key (we will refer to this block as the congruent IAT block) or respond more slowly when self-relevant words and introversion (and when non-self-relevant words and extraversion) require pressing the same key (further referred to as the incongruent IAT block). To fake *low* scores, Lisa could do the opposite (i.e., respond more slowly on the congruent IAT block or respond more quickly on the incongruent IAT block). Table 1 summarizes the faking strategies.

With the scoring algorithm suggested by Greenwald, Nosek, and Banaji (2003a, 2003b), an error penalty is added to the reaction times of every trial in which an error is committed. Hence, the strategic reduction or enhancement of errors affects the IAT score. Four possible strategies can be used to fake IAT scores by systematically manipulating errors (Table 1). When Hugh is trying to fake *high* scores, he could make an active effort to reduce errors on the congruent IAT block or increase errors on the incongruent IAT block. To fake *low* scores, Lisa could do the opposite (i.e., commit errors on the congruent IAT block and make an active effort to reduce errors on the incongruent IAT block).

1.3. Why people might use strategies beyond slowing down

Despite the multitude of faking strategies, past research has mainly focused on slowing down as an indication of faking. This shortcoming may be due to the assumption that slowing down is the *most likely faking strategy* (e.g., Cvencek, Greenwald, Brown, Gray, & Snowden, 2010; Kim, 2003). In fact, several studies have demonstrated that slowing down is a common strategy used by fakers (e.g., Cvencek et al., 2010; Fiedler & Bluemke, 2005). However, as shown above, slowing down is not the only strategy that fakers may come up with or that might be successful. Slowing down can sometimes be counterintuitive or even counterproductive, as we explain in the following.

Building on the assumption that fakers slow down, two recent approaches have suggested that fakers can be detected based on this supposed slowing down behavior. Cvencek et al. (2010) argued that slowing down on the block on which one was faster under nonfaking can be used as an indicator of faking (i.e., Combined Task Slowing; CTS; see Table 1). However, given the plausible scenario that a test taker was faster on the congruent block under nonfaking conditions and wants to fake higher scores, slowing down on that block would lead to lower instead of higher scores and would thus be counterproductive. Agosta, Ghirardi, Zogmaister, Castiello, and Sartori (2011) used a procedure that measures slowing down on the faster (i.e., congruent IAT block) compared to the single blocks (i.e., Ratio 150–10000 and Slow down 150–10000; see Table 1). However, if a test taker wants to fake higher scores, slowing down on that block will not lead to the desired outcome because the behavior produces lower scores as explained above.

The scenarios described above are examples of conditions in which *speeding up* is a strategy that would lead to success. Intuitively, one might think that speeding up would be impossible because participants are instructed to react as quickly as possible when completing the IAT under standard conditions, which makes further acceleration seem impossible. This view is backed by research that did not observe acceleration (e.g., Kim, 2003). However, some studies have indicated that acceleration is possible (Fiedler &

Table 1
Faking strategies and faking indices by IAT block and faking goal.

Block	Faking goal	
	Low scores	High scores
Conceptually derived faking strategies		
Congruent	*Slowing down on the congruent block* (difference in reaction time between the congruent IAT block under faking and the congruent IAT block at baseline)	*Acceleration on the congruent block* (difference in reaction time between the congruent IAT block at baseline and the congruent IAT block under faking)
	Increasing errors on the congruent block (difference in errors between the congruent IAT block under faking and the congruent IAT block at baseline)	*Reducing errors on the congruent block* (difference in errors between the congruent IAT block at baseline and the congruent IAT block under faking)
Incongruent	*Acceleration on the incongruent block* (difference in reaction time between the incongruent IAT block at baseline and the incongruent IAT block under faking)	*Slowing down on the incongruent block* (difference in reaction time between the incongruent IAT block under faking and the incongruent IAT block at baseline)
	Reducing errors on the incongruent block (difference in errors between the incongruent IAT block at baseline and the incongruent IAT block under faking)	*Increasing errors on the incongruent block* (difference between the incongruent IAT block under faking and the incongruent IAT block at baseline)
Previously suggested faking indices		
Faster and slower	*CTS* (difference between the slower IAT block under faking and the faster IAT block under nonfaking; Cvencek et al., 2010)	
Single and faster	*Ratio 150–10000* (ratio between the faster IAT block and the single IAT blocks under faking; Agosta et al., 2011)	
	Slow down 150–10000 (difference between the faster IAT block and the single IAT blocks; Agosta et al., 2011)	

Bluemke, 2005; Steffens, 2004). How can this discrepancy be explained? First, there might be practice effects (e.g., by becoming more familiar with the IAT stimuli and the procedure) that enable participants to react more quickly over time (Fiedler & Bluemke, 2005). If participants then strategically accelerate on particular blocks (e.g., on the congruent block when intending to fake high scores), this will lead to successful faking. Second, despite instructions to react as quickly as possible, participants might not push themselves to their limits under normal conditions and thus would be able to speed up when they are motivated to do so under faking. In addition to being possible, speeding up may even be likely. In fact, under certain conditions, speeding up might be an intuitive strategy to use. In many contexts (e.g., sports competitions or computer games), fast reactions are connected to higher scores. Therefore, participants might speed up especially when intending to fake high scores.

In addition to manipulating speed, fakers might manipulate errors to fake IAT scores. Naïve participants believe that not only reaction times but also errors are crucial to the IAT result (Steffens, 2004). As described above, errors can alter IAT scores (albeit indirectly, e.g., by penalties being added to reaction times). However, past research has provided inconsistent results. Some studies have reported no increase in error rates under faking (Cvencek et al., 2010; Fiedler & Bluemke, 2005), whereas others have found that participants tried to fake by committing errors (Steffens, 2004). In addition, past research has not analyzed whether alterations in error rates are connected to faking success. However, the strategic commission of errors is not the only faking strategy that can increase error rates; increases in error rates may also be a side effect of other faking attempts that require cognitive resources and thus impair correct responding.

1.4. Why faking direction matters

Past research has predominantly investigated only one faking direction within one study (i.e., faking low *or* high scores). This shortcoming might be the reason for some of the inconsistent findings because whether acceleration or slowing down is successful may depend on the faking direction. For example, acceleration on the congruent IAT block is necessary to fake high scores but slowing down on that block is necessary to fake low scores. Past research (Agosta et al., 2011; Cvencek et al., 2010) has proposed that slowing down could be used to detect and correct faked IAT scores, but these studies did not systematically take faking direction into account.

1.5. Why pre-existing knowledge matters

Results of past research on IAT faking are not comparable because some studies provided participants with hints about how to fake the IAT but others did not. When participants have background knowledge about faking strategies, they should be more likely to use those strategies. When they have no prior information, they may employ different strategies. Consequently, the study design must include a naïve condition and an informed faking condition. According to Steffens (2004), naïve participants believe that manipulating errors is important to the IAT result. Thus, one might expect that naïve fakers would be more likely to manipulate errors than participants who were informed about faking strategies that involve reaction times. In addition, naïve participants might intuitively slow down when intending to fake low scores and accelerate when intending to fake high scores (as this corresponds to rules that apply in various contexts such as sports competitions or computer games). By contrast, if participants are informed about alternative strategies for how to fake (e.g., slowing down on the incongruent block to fake high), they might opt for this less difficult strategy instead of effortful speeding up.

1.6. Why faking success matters

Until now, researchers have largely neglected the idea that participants might attempt to fake their IAT scores but might fail to alter their scores as desired. Assessing the extent to which strategies are connected to faking success is important because only successful fakers will distort the rank order of test scores. Not all faking strategies that people come up with will lead to successful faking. For example, a person might think that committing numerous errors will lead to a bad impression (i.e., low scores). However, committing errors in an unspecific way (i.e., on all IAT blocks) would not lead to the desired outcome. To fake low scores by means of manipulating errors, for example, errors must be committed on the congruent IAT block but not on the incongruent block. This example demonstrates the importance of investigating whether certain faking strategies are connected to faking success. In previous studies, this point was either not sufficiently considered (Cvencek et al., 2010) or faking success was defined as an inversion of the *D* score (Agosta et al., 2011). The latter does not seem appropriate because of the arbitrary metric and zero points of the IAT (Blanton & Jaccard, 2006) and the score distribution, which varies depending on the IAT used. Nevertheless, it is important to know whether a participant has attempted to fake but has failed or whether the attempt was successful.

1.7. The present study

We investigated the faking behavior of participants instructed to fake an extraversion IAT. The study comprised three experimental groups (faking high, faking low, and control) with three points of measurement (baseline, naïve faking, and informed faking). In the first step, ROC analyses (receiver operating characteristic curves; Green & Swets, 1966) were computed to investigate the faking strategies used by participants. In the second step, correlation analyses between faking strategies and faking success were computed because a strategy employed by fakers will not necessarily be successful in altering the IAT result as desired. In detail, our study investigated the following questions and hypotheses:

1. To understand how people fake the IAT under different faking conditions, we investigated several faking strategies (slowing down, acceleration, and manipulation of errors) that can be used by fakers to alter their IAT scores. We investigated whether fakers used the following faking strategies under faking relative to nonfaking at baseline: (a) slowing down on the congruent block and on the incongruent block, (b) accelerating on the congruent block and on the incongruent block, (c) increasing errors on the congruent block and on the incongruent block, and (d) reducing errors on the congruent block and the incongruent block. We also investigated whether participants who fake stand out because of unspecific slowing down behavior as suggested previously (Agosta et al., 2011; Cvencek et al., 2010).
2. We tested whether the various faking strategies were correlated with faking success. At first glance, it might appear circular to let participants fake and then test whether faking strategies are connected to faking success. Actually, it is not. Note that at first, participants did not receive any hints about how to achieve their faking goals. Furthermore, it is possible that some strategies will have side effects that will render them ineffective for achieving the desired goal (e.g., someone might try to speed up on the congruent block to fake a stronger association but this could increase the number of errors if the person's speed exceeds the person's ability).

3. Based on our assumption that fakers use different strategies depending on the faking direction (i.e., low or high scores), we asked participants to fake *high* scores as well as *low* scores. We expected that participants attempting to fake low scores might intuitively slow down, whereas participants attempting to fake high scores would accelerate.
4. We expected that pre-existing knowledge about how to fake the IAT would matter with respect to the choice of faking strategy. Thus, participants were first asked to fake, but they did not receive a strategy for how to fake. They then received information about the IAT's rationale and were asked to fake again. We expected that naïve participants would manipulate errors and reaction times, whereas participants informed about faking strategies that rely on reaction times would make use of those strategies and predominantly try to alter their reaction times. Additionally, we expected that naïve participants might intuitively slow down when intending to fake low scores and accelerate when intending to fake high scores. When given faking strategies, participants were expected to opt for the less difficult strategy of slowing down instead of effortful speeding up.

2. Method

2.1. Participants

The sample consisted of 84 participants (64 females; 74 students) with an average age of 22.4 years (*SD* = 4.5). Data were collected from volunteers in exchange for personal feedback and partial course credit.

2.2. Procedure

Participants were assigned to one of three conditions: (a) control group, (b) faking condition LH (faking low scores first and then faking high scores), (c) faking condition HL (faking high scores first and then faking low scores). Participants completed an extraversion IAT three times. In Session 1, all participants completed the IAT once without faking instructions (baseline assessment). Participants in the control group then completed the IATs repeatedly without further instructions. We asked participants in the faking conditions to imagine being in a personnel selection scenario and to fake either high extraversion (to maximize their chances of being offered an attractive job) or low extraversion (to avoid being offered an unattractive job) on the IAT. They were not provided with information about the IAT's rationale or faking strategies (*naïve faking*). In Session 2, participants in the faking conditions were told how to fake the IAT (*informed faking*). They were given information about differences between congruent and incongruent IAT blocks. The instructions also included information about the computation of the IAT score based on differences in reaction times between the two critical blocks. Participants were provided with a recommended strategy that included slow reactions on the congruent block and fast reactions on the incongruent block to fake low scores. The instructions for faking high scores were the other way around. To minimize the influence of practice effects, participants had to fake high scores if they had previously faked low scores (faking condition LH) or low scores if they had previously faked high scores (faking condition HL).

2.3. Extraversion IAT

We used an extraversion IAT (Back, Schmukle, & Egloff, 2009) as described in the example in Section 1.1. The single dimension Practice Blocks 1, 2, and 5 each included 24 trials. The CombinedBlocks 3, 4, 6, and 7 each included 48 trials. Data from the combined blocks were used to compute IAT scores (D_1 measure; Greenwald, Nosek, & Banaji, 2003a, 2003b). Extremely long responses (i.e., more than 10,000 ms) were dropped. In contrast to the example (see Section 1.1), our IAT was counterbalanced for combined block order between participants (i.e., half of the participants first worked on the congruent IAT block and then on the incongruent IAT block; the other half of the participants first worked on the incongruent IAT block and then on the congruent one). The presentation of critical blocks was held constant within participants. Split-half reliabilities were good to excellent at all points of measurement (Table 2).

Table 2
Split-half reliabilities, descriptive variables, and post hoc comparisons regarding the *D* scores of the extraversion IAT.

Measurement occasion	Split-half reliabilities	Experimental group		
		Faking condition LH *M* (*SD*)	Faking condition HL *M* (*SD*)	Control group *M* (*SD*)
1 (Baseline)	.86	.27 (.46)$_{a1}$	.09 (.47)$_{a1}$	.36 (.38)$_{a1}$
2 (Naïve faking/retest)	.91	−.23 (.56)$_{a2}$	.41 (.55)$_{b2}$	.31 (.43)$_{b1}$
3 (Informed faking/retest)	.97	1.04 (.25)$_{a3}$	−.79 (.68)$_{b3}$	.33 (.37)$_{c1}$

Note: *N* = 84; different alphabetic subscripts indicate significant differences between experimental groups (i.e., columns); different numeric subscripts identify significant differences between measurement occasions (i.e., rows) at $p < .05$.

2.4. Analytic strategy

We conducted two kinds of analyses to investigate which strategies fakers use to fake the IAT and whether those strategies are successful in altering IAT scores. First, in order to examine which strategies were employed by fakers, we used ROC analyses to investigate how well each of the strategies was able to predict whether participants belonged to the control group or a faking group (cf. Cvencek et al., 2010). If most of the fakers use a certain strategy (e.g., slowing down on the congruent IAT block to fake low scores), then it should be possible to distinguish fakers from nonfakers on the basis of the respective behavior. In ROC analyses, hit rates (for successfully identifying individuals in the faking condition) are plotted as a function of false-alarm rates (falsely identifying participants in the control group as fakers). The area under the curve (*AUC*) shows the success of each strategy in predicting whether a participant belonged to the faking group or the control group (i.e., whether the faking status could be assigned at levels above chance). If the *AUC* differed significantly from the chance rate of .50, the strategy was typically used by fakers in contrast to nonfakers at levels above chance. Analyses were performed separately for the faking high and faking low conditions (compared to the control group in each case).

Second, we ran correlation analyses to investigate the extent to which the strategies were connected to *faking success*. This is an important aspect to investigate because it is not necessarily the case that an employed strategy will also succeed in altering the IAT scores as desired. Faking success was computed as the difference in *D* scores between a faked and a nonfaked IAT (*D* change; cf. Cvencek et al., 2010). Faking success for participants faking *low* scores was calculated by subtracting the *D* score of the faked IAT from the *D* score of the nonfaked IAT completed at baseline assessment. Faking success for participants faking *high* scores was calculated by subtracting the *D* score of the nonfaked IAT from the *D* score of the faked IAT. Thus, positive values indicate successful faking independent of the specific faking direction. We also computed *D* change for the control group (here, as the difference between the baseline assessment and repeated measurements to represent unspecific changes in *D* scores that were not due to the faking instructions). We then computed correlation analyses between *D* change and the respective faking strategy to assess the extent to which the usage of a strategy was linked to success in changing the score. We used Fisher's *z* test (Fisher, 1950) to compare the correlations in the faking groups to those in the control group to determine whether the correlations differed from unspecific effects. Because of relatively low power as revealed by a power analysis using G*power (Faul, Erdfelder, Buchner, & Lang, 2009), we defined $p < .10$ as significant in the Fisher's *z* tests.

For reasons of comparison, we additionally conducted both ROC analyses and correlation analyses with Agosta et al.'s (2011) and Cvencek et al.'s (2010) indices that are intended to measure strategic slowing down in the IAT.

3. Results

3.1. Manipulation check

To examine whether the IAT could be faked successfully, we conducted a 3 (measurement occasion) 3 (experimental group) ANOVA with repeated measures on IAT *D* scores. The main effect of measurement occasion was not significant, $F(1.64, 133.02) = 0.53$, *ns*, $\eta^2_{\text{partial}} = .01$, $\omega^2 = .00$. The significant main effect of group, $F(2, 81) = 21.44$, $p < .05$, $\eta^2_{\text{partial}} = .35$, $\omega^2 = .33$, was qualified by the expected significant and large interaction effect, $F(3.28, 133.02) = 52.41$, $p < .05$, $\eta^2_{\text{partial}} = .56$, $\omega^2 = .67$.[2] Participants instructed to fake high had significantly higher scores than at baseline, and participants who faked low had significantly lower scores, indicating successful faking. There were no significant differences between the experimental groups at baseline or between measurements in the control group (Table 2).

3.2. Distribution of error rates across the different faking conditions

We conducted a 3 (measurement occasion) 3 (experimental group) ANOVA with repeated measures on error rates. The analysis yielded significant main effects of measurement occasion, $F(1.18, 94.52) = 35.31$, $p < .05$, $\eta^2_{\text{partial}} = .30$, and group, $F(2, 81) = 21.08$, $p < .05$, $\eta^2_{\text{partial}} = .35$, and a significant interaction effect, $F(2.33, 94.52) = 31.18$, $p < .05$, $\eta^2_{\text{partial}} = .44$. No significant differences were found between the experimental groups at baseline or between measurements in the control group (Table 3). Faking did not result in generally higher error rates. Error rates were significantly higher only for naïve faking of low scores compared to baseline and also compared to the error rates of the two other groups (i.e., naïve faking of high scores and control).

Table 3
Descriptive variables and post hoc comparisons regarding the error rates of the extraversion IAT.

Measurement occasion	Experimental group		
	Faking condition LH *M* (*SD*)	Faking condition HL *M* (*SD*)	Control group *M* (*SD*)
1 (Baseline)	5.10 (3.89)$_{a1}$	4.61 (4.17)$_{a1}$	4.77 (3.44)$_{a1}$
2 (Naïve faking/retest)	34.12 (24.89)$_{a2}$	6.90 (6.06)$_{b1}$	5.79 (4.22)$_{b1}$
3 (Informed faking/retest)	7.40 (6.32)$_{a3}$	6.50 (5.05)$_{a1}$	5.99 (4.40)$_{a1}$

Note: $N = 84$; different alphabetic subscripts indicate significant differences between experimental groups (i.e., columns); different numeric subscripts identify significant differences between measurement occasions (i.e., rows) at $p < .05$.

3.3. What fakers do at naïve faking and what leads to faking success

3.3.1. Faking low

Recall Lisa who wanted to fake low scores. Potentially successful strategies would be to react more slowly on the congruent IAT block, react more quickly on the incongruent block, increase the number of errors on the congruent block, and reduce the number of errors on the incongruent block.

ROC analyses (Table 4 and Fig. 1) showed that naïve participants who faked low scores could be successfully distinguished from nonfakers with the help of the strategies to slow down and to increase errors on the congruent block (*AUCs* = .89 and .96, respectively), but not via the strategies to accelerate or to reduce errors on the incongruent block (both *AUCs* = .31). The correlation analyses (see Table 4) between the strategies and faking success revealed that slowing down on the congruent block, acceleration on the incongruent block, and reducing errors on the incongruent block were positively correlated with faking success. However, Fisher's *z* tests revealed that only slowing down on the congruent block was more strongly correlated with faking success in the faking (low) group than with unspecific changes in the *D* score in the control group, $z = 2.48$ ($p < .10$). By contrast, acceleration and error reduction on the incongruent block were also associated with *D* change in the control group, which may be due to practice effects (cf. Fiedler & Bluemke, 2005).

According to Cvencek et al. (2010), Lisa's faking attempts should be indicated by her slower reaction times on the combined block on which she had reacted more quickly under nonfaking (i.e., as represented by the CTS index). According to Agosta et al. (2011), Lisa's faking attempts would stand out by her slower reaction times on the congruent IAT block compared to single blocks (i.e., as represented by the indices Ratio 150–10000 and Slow down 150–10000). CTS correctly classified participants as belonging to the control or faking (low) group, but Ratio 150–10000 and Slow down 150–10000 could not classify faking. None of the faking indices was significantly positively connected with faking success (Table 4), demonstrating that the behavior measured by these indices does not lead to faking success or that the behavior is even counterproductive.

3.3.2. Faking high

Recall Hugh, who wanted to fake high scores. Potentially successful strategies are to react more slowly on the incongruent IAT block, react more quickly on the congruent IAT block, increase the number of errors on the incongruent IAT block, and reduce the number of errors on the congruent IAT block.

In the ROC analyses (Table 4 and Fig. 1), participants could be classified correctly as belonging to the faking (high) group or control group *only* via the strategy to accelerate on the congruent block. Acceleration on the congruent block was also the only strategy that was significantly positively correlated with faking success (Table 4) and that was more strongly correlated with faking success in the faking (high) group than with unspecific changes in the control group (Fisher's $z = 3.75$, $p < .10$).

According to Agosta et al. (2011) and Cvencek et al. (2010), Hugh's faking attempts would be indicated by the same faking behavior as Lisa's. The faking indices developed by Cvencek et al. (2010) and Agosta et al. (2011) however, were unable to correctly classify participants in the faking (high) and control groups. The indices suggested by Agosta et al. (2011) were actually negatively connected to faking success, thus suggesting that the behavior measured by those indices would have a negative impact on faking success.

3.4. What fakers do at informed faking and what leads to faking success

3.4.1. Faking low

Recall that Lisa can use the following strategies to fake low scores: react more slowly on the congruent IAT block, react more

[2] The *df* of all ANOVAs were adjusted according to Greenhouse–Geisser (Weiner, Schinka, & Velicer, 2012).

Table 4
Implementation and success of faking strategies and indices depending on faking direction and faking instructions.

Faking strategies and indices	Naïve faking				Informed faking			
	Implementation		Correlation with faking success	Correlation with unsystematic changes	Implementation		Correlation with faking success	Correlation with unsystematic changes
	AUC	*SE*	*r*	*r*	*AUC*	*SE*	*r*	*r*
Faking low								
Slowing down on the congruent block	**.89**	.06	**.52**	−.13	**.78**	.07	.51	.39
Acceleration on the incongruent block	.31	.08	.48	.80	.61	.08	**.70**	.10
Increasing errors on the congruent block	**.96**	.03	−.28	.16	.60	.08	−.10	.28
Reducing errors on the incongruent block	.31	.08	.59	.38	.32	.08	−.26	.35
CTS	**.89**	.05	.14	−.31	**.73**	.07	−.12	.14
Ratio 150–10000	.24	.07	−.16	−.01	.52	.08	−.28	.10
Slow down 150–10000	.19	.06	−.12	−.05	.52	.08	−.32	.10
Faking high								
Slowing down on the incongruent block	.63	.08	.73	.80	**.82**	.06	**.55**	.10
Acceleration on the congruent block	**.73**	.07	**.73**	−.13	.61	.08	.53	.39
Increasing errors on the incongruent block	.55	.08	.43	.38	.58	.08	.28	.35
Reducing errors on the congruent block	.64	.08	.02	.16	.48	.08	−.09	.28
CTS	.55	.08	.47	.31	**.81**	.06	−.18	−.14
Ratio 150–10000	.42	.08	−.43	.01	.43	.08	.11	−.10
Slow down 150–10000	.42	.08	−.52	.05	.43	.08	.10	−.10

Note: *AUCs* in bold indicate that the strategy or index classified participants as belonging to the control or faking group at levels above chance (>.50). Faking success = changes in IAT scores according to faking instructions. Unsystematic changes = changes in IAT scores in the control group (i.e., not due to faking instructions). Correlations $|r| \geqslant .38$ were significant at $p < .05$, correlations $|r| \geqslant .48$ were significant at $p < .01$. Correlations printed in bold indicate that the significant positive correlation between the relevant faking strategy and *D* change (i.e., faking success) in the faking group was significantly higher than the correlation between the respective behavior and *D* change in the control group according to Fisher's *z* tests at $p < .10$.

quickly on the incongruent block, increase the number of errors on the congruent block, and reduce the number of errors on the incongruent block.

ROC analyses (Table 4 and Fig. 1) showed that informed fakers used *only* the strategy of slowing down on the congruent IAT block to fake low scores. The correlation analysis (Table 4) revealed that

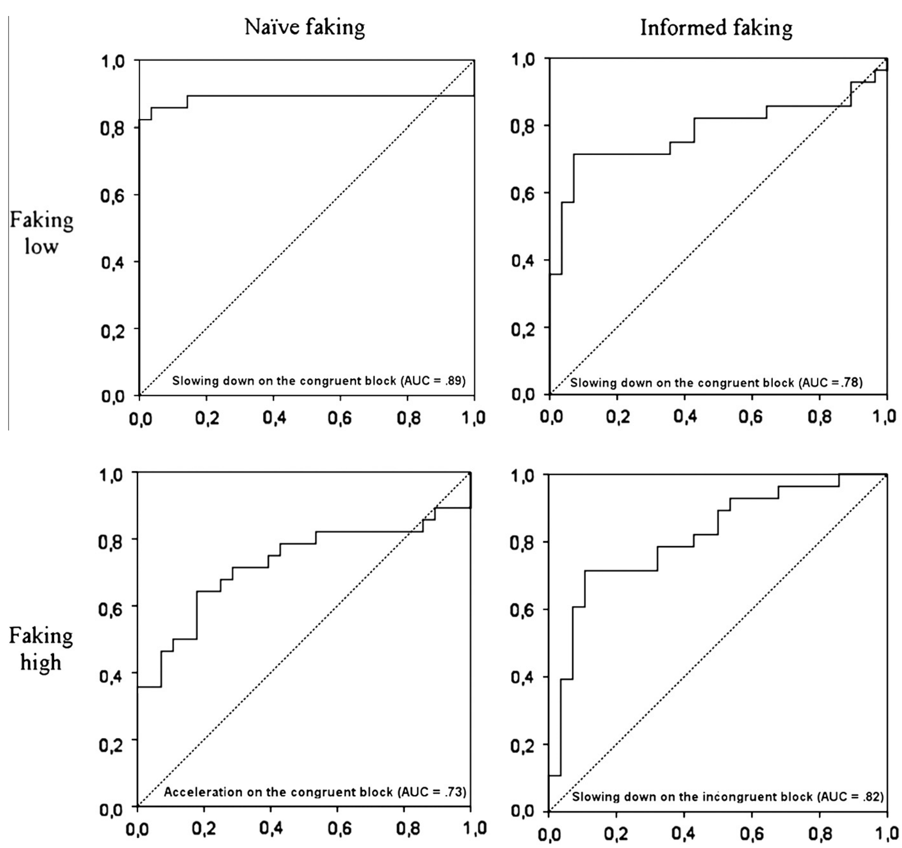

Fig. 1. ROC analyses (Green & Swets, 1966) for the conceptually developed faking strategies participants use to fake the IAT in the desired direction. The hit rate (proportion of correctly identified faking participants) is plotted on the *y*-axis against the false alarm rate (proportion of nonfaking participants incorrectly identified as fakers) on the *x*-axis. The diagonal line represents chance success. The area under the curve (*AUC*) corresponds to the percentage correct on a two-alternative forced-choice detection task.

slowing down on the congruent block was positively correlated with faking success for participants attempting to fake low scores. Due to low power, however, the correlation between slowing down on the congruent block and faking success was descriptively but not significantly higher than the correlation between slowing down on the congruent block and unspecific changes in the control group, Fisher's $z = .53$ (*ns*).

Although a ROC analysis had shown that the strategy to accelerate on the incongruent block was not typically used by fakers in this condition, the strategy was significantly positively connected to faking success, indicating that the few fakers who used the strategy were successful at altering their scores as desired. This correlation differed significantly from the correlation with unspecific changes for participants in the control group, Fisher's $z = 2.73$ ($p < .10$).

Regarding the indices, CTS (Cvencek et al., 2010) correctly classified participants in the faking (low) and control groups. Ratio 150–10000 and Slow down 150–10000 (Agosta et al., 2011) could not assign faking status. None of those faking indices showed a significant positive relation to faking success, and some of the (nonsignificant) relations were actually negative (Table 4).

3.4.2. Faking high

Recall that Hugh can use the following strategies to fake high scores: react more slowly on the incongruent IAT block, react more quickly on the congruent IAT block, increase the number of errors on the incongruent IAT block, and reduce the number of errors on the congruent IAT block.

ROC analyses (Table 4 and Fig. 1) showed that participants could be correctly classified as belonging to the faking (high) or control group *only* via the strategy to slow down on the incongruent block. This strategy was also positively connected with faking success (Table 4), and its correlation with faking success was significantly higher than its correlation with unspecific changes in the control group, Fisher's $z = 1.80$ ($p < .10$).

Regarding the indices, CTS (Cvencek et al., 2010) correctly classified participants in the faking (high) and control groups. However, CTS was not significantly connected with faking success. Ratio 150–10000 and Slow down 150–10000 (Agosta et al., 2011) could not assign faking status and were not correlated with faking success either.[3]

4. Discussion

In the present study, we investigated (a) which of the conceptually possible faking strategies are in fact used and whether faking direction (i.e., high or low scores) and pre-existing knowledge on how to fake (i.e., naïve vs. informed faking) have an impact on the choice of faking strategies, (b) which faking strategies are connected to faking success, and (c) whether the indices assumed to detect the typical slowing down behaviors of fakers (Agosta et al., 2011; Cvencek et al., 2010) are connected to faking success and can be used to detect fakers.

4.1. What fakers do to fake the IAT

Our results highlight a very important issue: As hypothesized, participants use different faking strategies depending on the desired faking direction (high or low) and on their pre-existing knowledge about how to fake the IAT.

Naïve participants faking low scores used the strategy of slowing down on the congruent block. This result is in line with earlier accounts that suggested that participants tend to slow down to fake the IAT (e.g., Cvencek et al., 2010), but our finding is more specific because it shows that slowing down occurs on the *congruent block*. Naïve participants faking low scores also showed an increase in the number of error rates on the congruent block. This result dovetails with findings that have indicated that participants believe that errors are important when faking IAT scores (Steffens, 2004), and it demonstrates that they use other strategies in addition to slowing down. Although reducing errors or acceleration on the incongruent block would also have been possible faking strategies, they were not used, probably because they are more difficult to employ; thus, we can conclude that participants apparently opt for the easier options.

Naïve participants faking high scores accelerated their reaction times on the congruent block, which indicates that, as expected, faking strategies depend on faking direction. This result also provides further evidence that slowing down is not the only strategy used by fakers. It is noteworthy that naïve fakers sped up on the congruent IAT block instead of slowing down on the incongruent one, which would be the easier strategy. Apparently, manipulating the congruent block is the more intuitive strategy. Manipulating the incongruent block requires participants to understand the rationale behind the IAT and conclude that they will appear more extraverted if they present themselves as less introverted (i.e., if they slow down on the self-introversion blocks). It is also noteworthy that speeding up was actually possible in the present study, given that several studies had not observed acceleration (e.g., Kim, 2003). Conventional wisdom says that acceleration is not possible given that the standard IAT instructions already ask participants to react as quickly as possible. Our results suggest that either practice effects or motivational effects most likely play a role.

Informed participants faking low scores used the strategy of slowing down on the congruent block but did not accelerate on the incongruent IAT block. Thus, in terms of reaction times, the strategy they used was the same strategy that was used to fake low in the naïve condition. This result is not surprising because slowing down on the congruent block seems to be the easiest and most intuitive faking strategy. In contrast to naïve participants, however, informed participants in the faking-low condition did not manipulate their errors. This indicates that participants who know that reaction times are sufficient for faking will favor the strategy of influencing reaction times.

Informed participants faking high scores slowed down on the incongruent IAT block. They did not accelerate on the congruent IAT block (as naïve fakers did when faking high). After they had learned that slowing down on the incongruent IAT block will produce the same result as accelerating on the congruent IAT block, they probably chose to slow down because it required less effort. Apparently, informed fakers understood the test's rationale and specifically chose to slow down on the relevant block (i.e., on the congruent block to fake low scores and on the incongruent block to fake high scores) instead of using the naïve strategy of manipulating the congruent block only.

4.2. What leads to faking success?

We tested which faking strategies were connected with faking success. All reaction-time-related strategies that we had derived

[3] We conducted the same analyses for the other indications of faking suggested by Cvencek et al. (2010). For naïve faking of low scores, the indices a, b, c, and e (Cvencek et al., 2010) correctly identified participants as belonging to the control or faking groups. However, none of the behaviors measured by the indices was significantly positively correlated with faking success. For naïve faking of high scores, the indices could not be used to correctly classify participants in the faking and control groups. For informed faking of low scores, the behaviors measured by the indices a, c, and e but not index b identified participants in the faking and control groups. Yet, none of the behaviors measured by these indices were significantly positively related to faking success. For informed faking of high scores, the indices a, c, and e, but not index b correctly identified participants as belonging to the faking and control groups. Still, none of the behaviors measured by these indices was significantly positively related to faking success. Accordingly, the indices cannot be used to detect successful faking. Detailed results can be obtained from the corresponding author upon request.

theoretically and that participants used in their faking attempts were also connected to faking success. By contrast, none of the error-related strategies were connected to faking success, although participants who faked low scores in the naïve condition stood out because of increased error rates. The fact that those increased error rates were not related to faking success indicates that these participants did not increase their errors on the congruent IAT block specifically but rather increased their errors across all blocks. Participants may have thought that committing errors in general would lead to lower scores, but in fact, this strategy fails to alter the IAT score as desired. Errors may also be a side effect of faking attempts that require cognitive resources and thus impair correct responding. However, the latter seems unlikely as higher error rates occurred only in this particular faking condition (i.e., naïve faking of low scores).

4.3. Is it sufficient to look at slowing down to reveal successful faking?

The IAT faking indices suggested by Agosta et al. (2011) and Cvencek et al. (2010) are based only on slowing down behavior. However, as our results clearly show, participants use different faking strategies (including acceleration) that depend on faking direction and pre-existing knowledge about the IAT and how to fake it. Therefore, detecting fakers with these published indices must fail in many conditions. Accordingly, the CTS index by Cvencek et al. (2010) worked only in some conditions (i.e., when faking low scores or after participants had been informed about faking strategies). The indices by Agosta et al. (2011) were not at all successful in revealing faked IAT scores in our study. More importantly, the slowing down behaviors as represented by the published indices were not (and in some cases even negatively) related to faking success. Therefore, although partly successful at indicating *faking attempts*, the index by Cvencek et al. (2010) is not able to reveal *successful faking*. How can this discrepancy between faking attempts and successful faking be explained? Participants might possibly become slightly slower as they think about how they can fake (i.e., slowing down might represent some kind of side effect but not necessarily a behavior that is strategically employed). This slowing down may stand out most on the IAT block on which participants had reacted more quickly under nonfaking and can thus be used to detect faking attempts in some but not all conditions.

4.4. Limitations

Our study is limited because we used only an extraversion IAT. Future studies should replicate the findings using IATs that measure other constructs. Another limitation is the small sample that led to low power in some of our analyses. Due to the low power of our correlation analyses and Fisher's z tests, our tests might have been too conservative despite the fact that we adjusted the significance level of Fisher's z tests. Future studies should thus aim to replicate and extend the findings with larger samples.

In addition, our sample predominantly consisted of first-year students, who in most instances had never completed an IAT before. Future research should use a more diverse sample to investigate whether prior experience with the IAT influences faking success or the choice of faking strategy. Furthermore, we used difference scores to measure faking success. Despite the long tradition and widespread use of difference scores, statistical tests that use difference scores are assumed to suffer from methodological problems (e.g., lower reliability, conservative testing; cf. Edwards, 2001). Notwithstanding these problems, we opted for difference scores in the present study given that we wanted to compare our results to the approaches suggested in past studies that also relied on difference scores. The fact that we found significant effects despite the potential for low reliability shows that these effects seem to be robust. Still, care has to be taken when interpreting nonsignificant findings. However, given that the reliabilities were comparable across conditions, it is unlikely that low reliabilities would account for all of our nonsignificant findings.

Finally, there was a certain overlap between the predictor and criterion variables (because they are both computed from reaction times on the IAT), and this overlap has the potential to lead to confounding effects. Nevertheless, we used this approach because we considered it useful to analyze our data in a manner that was comparable to previous studies (cf. Cvencek et al., 2010) and because the correlations between the variables differed across several faking conditions, thus rendering a strong confounding effect less likely.

4.5. Conclusions and practical implications

The results of our study provide important insight into how participants fake under different conditions. Our results have theoretical implications for the existing assumptions about how people fake the IAT and may, together with further evidence, eventually have practical implications for the ability to detect fakers.

First, the assumption of slowing down as the *most likely faking strategy* (e.g., Cvencek et al., 2010; Kim, 2003) does not apply to both faking directions. Acceleration is possible and leads to faking success when faking high scores. However, participants try to fake via acceleration only when they are not informed about a less effortful faking strategy (i.e., slowing down on the other block). Thus, pre-existing knowledge matters, as participants typically seem to opt for the easiest strategy available. Second, successful naïve fakers tend to manipulate the congruent IAT block (i.e., by accelerating to fake high scores and by slowing down to fake low scores). Apparently, naïve faking is less about slowing down or speeding up per se but more about specifically manipulating reaction times on the *congruent* block but not the incongruent one—probably because the former is more intuitive. Third, in some conditions, faking attempts stand out because of increases in error rates, but committing errors is not a successful faking strategy. Consequently, errors cannot be used to detect successful fakers.

With respect to practical implications, we recommend that researchers use great caution when implementing faking indices to detect fakers or to correct IAT scores. Researchers must consider that some individuals in a dataset will be naïve fakers (i.e., they spontaneously develop a faking strategy), whereas others will know about faking strategies (e.g., from the TV or the internet). Thus, they may use different faking strategies even when trying to fake the same impression (e.g., high scores). Moreover, depending on the construct and the goal, some people may try to fake high scores, whereas others may try to fake low scores, and they will use different faking strategies accordingly. Last but not least, some people will try to fake but others will not, and not all faking attempts will be successful. Given that the rank order of test takers will change only if participants fake successfully, it is important to differentiate between attempted and successful faking. Published faking indices based solely on slowing down behavior (Agosta et al., 2011; Cvencek et al., 2010) were not connected to faking success in most conditions and thus cannot necessarily be used to detect successful fakers. However, when using the IAT for selection processes or when attempting to correct faked scores as Cvencek et al. (2010) suggested, it is important to correct only for IAT scores that are actually affected by faking. Therefore, more work must be performed before indices to detect faked IAT scores or algorithms to correct faked IAT scores can be used without reservation.

Acknowledgment

The authors thank Dario Cvencek for providing detailed information on the methods used to analyze faking in his study.

References

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (2011). Detecting fakers of the autobiographical IAT. *Applied Cognitive Psychology, 25*, 299–306. http://dx.doi.org/10.1001/acp. 1691.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*, 533–548. http://dx.doi.org/10.1037/a0016229.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27–41. http://dx.doi.org/10.1037/0003-066X.61.1.27.

Cvencek, D., Greenwald, A. G., Brown, A. S., Gray, N. S., & Snowden, R. J. (2010). Faking of the Implicit Association Test is statistically detectable and partly correctable. *Basic and Applied Social Psychology, 32*, 302–314. http://dx.doi.org/10.1080/01973533.2010.519236.

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43*, 972–978. http://dx.doi.org/10.1016/j.jesp. 2006.10.007.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265–287. http://dx.doi.org/10.1177/109442810143005.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology, 27*, 307–316. http://dx.doi.org/10.1207/s15324834basp2704_3.

Fisher, R. A. (1950). *Statistical methods for research workers* (11th ed.). London: Oliver & Boyd.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464.

Greenwald, A., Nosek, B., & Banaji, M. (2003a). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216. http://dx.doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A., Nosek, B., & Banaji, M. (2003b). 'Understanding and using the Implicit Association Test: I. An improved scoring algorithm': Correction to Greenwald et al. (2003). *Journal of Personality and Social Psychology, 85*. http://dx.doi.org/10.1037/h0087889.

Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly, 66*, 83–96. http://dx.doi.org/10.2307/3090143.

McDaniel, M., Beier, M., Perkins, A., Goggin, S., & Frankel, B. (2009). An assessment of the fakeability of self-report and implicit personality measures. *Journal of Research in Personality, 43*, 682–685. http://dx.doi.org/10.1016/j.jrp. 2009.01.011.

Röhner, J., Schröder-Abé, M., & Schütz, A. (2011). Exaggeration is harder than understatement, but practice makes perfect! Faking success in the IAT. *Experimental Psychology, 58*, 464–472. http://dx.doi.org/10.1027/1618-3169/a000114.

Steffens, M. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*, 165–179. http://dx.doi.org/10.1027/1618-3169.51.3.165.

Weiner, I. B., Schinka, J. A., & Velicer, W. F. (2012). *Handbook of psychology. Research methods in psychology* (Vol. 2). New Jersey: Wiley.