

Exercise effects in the Implicit Association Test (IAT)

Abstract

Greenwald, McGhee and Schwarz (1998a) assume that individual differences in implicit cognition can be measured by means of the Implicit Association Test (IAT). The IAT has been criticised by some authors contenting that the IAT effect does not reflect implicit cognition alone but is also susceptible to other influences, e.g. "task-switch costs". A modified version of the IAT is suggested here which allows testing the influence of exercise effects and task-switch effects. In two experiments participants completed a modified version of the IAT. In the first study we used the domain of "aggression", and in the second study we used the "self concept" domain to measure implicit attitudes.

The results confirm the hypothesis, that the IAT effect can not be traced back to implicit attitudes alone. It is shown that the IAT effect is influenced by simple exercise effects but not so much by task switch-costs.

Keywords: implicit attitudes, gender stereotypes, IAT, self-evaluation, exercise effects, implicit

Besides other response latency procedures (e.g. Fazio et al., 1986), the IAT (Implicit Association Test) by Greenwald et al. (1998 a) was proposed as a method to assess implicit attitudes. In this task, items from bipolar target concepts like flowers vs. insects and bipolar attributes like bitter vs. nice are presented in a forced-choice procedure, and the strength of the association between target concepts and associated attributes is measured. To understand the procedure it is necessary to distinguish an evaluatively compatible and an evaluatively incompatible condition. In the evaluatively compatible condition the target concept and the associated attribute of the same valence are mapped to the same response key (e.g. flowers and nice to the left key). In the evaluatively incompatible condition the target concept shares the same response key with an attribute of a different valence (e.g. insect – nice).

The so called IAT effect is defined as the difference in reaction time between the evaluatively incompatible and the evaluatively compatible condition. The IAT effect is then taken as a measure for the implicit association between the target concept and its associated attribute.

Since the beginning of its publication in 1998 the Implicit Association Test (Greenwald et al., 1998 a) has been discussed controversially. Nevertheless there is a vast literature supporting the IAT (Rudman, Greenwald, Mellott & Black, 1999; Greenwald & Farnham, 2000; Banse, Seise, & Zerbes, 2001; Cunningham, Preacher & Banaji, 2001; McConnell & Leibold, 2001). However, critical studies or alternative explanations for the IAT effect can be found in some recent publications (Rothermund & Wentura, 2001; Mierke & Klauer, 2001; De Houwer, 2001).

In particular, Mierke and Klauer (2001) specified a phenomenon, which was already mentioned by Greenwald et al. (1998a). They pointed out that the strength of the IAT effect can not be exclusively explained by the underlying attitude, and that a part of the reaction time difference can be traced back to different costs when changing among the tasks (from compatible to incompatible versus incompatible to compatible). To perform the given tasks in the proper way, response-incongruent reactions have to be inhibited. That means that costs arise when a response, which was practiced before, is disrupted by switching to a different task. In this case a change from a compatible to an incompatible condition would be more difficult than a change from an incompatible to a compatible condition. Therefore, by changing from the compatible to the incompatible task a greater IAT effect would result than by changing from the incompatible to the compatible task. If the difference between the conditions only depended on the implicit attitude, the order of the presentation of the

evaluatively compatible and incompatible condition should not have an influence on the size of the IAT effect. If the IAT effect contained switch-costs, the effect-size is not exclusively founded on evaluative associations between the target and the attribute categories.

In our opinion it is not clear, whether the effect obtained is based on simple exercise or on so called switch-costs. We assume that complex discrimination tasks such as those in the IAT lead to greater exercise effects than simple tasks. In this case a transition from the compatible to the incompatible task in the IAT corresponds to a change from an easy to a difficult task. Therefore, the so-called switch-costs postulated by Mierke and Klauer (2001) could also be described as simple exercise effects. A demanding exercise, however, should result in a high release of additional cognitive resources. If only fewer resources are needed for the following task, this would lead to a facilitation and therefore to shorter reaction times.

In addition to the switch-costs postulated by Mierke and Klauer (2001), Rothermund and Wentura (2001) assumed a further influence on the IAT effect. They could show that "Figure-Ground Asymmetries" could account for the IAT effect and suggest that these effects do not rely exclusively on evaluative associations between the target and the attribute categories.

In consideration of these critical objections there is no doubt that the IAT is not independent of other influences. Even Greenwald and Nosek (2001, S.7) admit that: "there is, accordingly, no doubt that IAT measures are susceptible to influence by a variety of independent variable manipulations." Predominantly the switch – costs, postulated by Mierke and Klauer (2001), and the figure – ground asymmetries, shown by Rothermund and Wentura (2001), seem to have an important influence on the IAT.

The present research will consist of two studies on implicit stereotyping comparing the influence of possible switch-costs (Mierke & Klauer, 2001) or exercise-effects on the IAT: the measurement of implicit aggression in experiment 1 and implicit self-concept in experiment 2. Instead of the usual five block version, we used a modified version of the IAT with a sequence of seven experimental blocks by repeating each the compatible and the incompatible blocks as an additional experimental condition.

Study 1

In the following study we tested the assumption that part of the IAT – effect can be traced back to simple exercise effects. A modified version of the IAT is used to assess the precision of the IAT measures for implicit gender stereotypes.

The following hypotheses are investigated:

1. Aggressive behaviour is more strongly associated with men than with women. (IAT effect)
2. The magnitude of this IAT effect is equal for male and female subjects. (IAT effect and sex)
- 3a. A reduction of the IAT effect due to the influence of exercise will result when the compatible and the incompatible conditions are repeated. (IAT effect and exercise)
- 3b. Familiarity with the test procedure is expected to reduce the IAT effect on replication of the IAT as a whole.
4. The IAT effects are stronger when the compatible condition is presented first (task-switching hypothesis: IAT-effect and sequence of IAT conditions).

Method

Material

Stimulus material of words with clearly identified "aggressive" or "peaceful" contents had to be selected for the experiments. Therefore, a pool of 80 words (40 verbs with an "aggressive" contents and 40 with a "peaceful" content) was collected first. 86 participants (52 female and 34 male) aged 16 to 78 years ($M = 37.98$, $SD = 15.10$) were asked to rate the stimulus words by using a seven-point "Likert"- scale from -3 to 3. A cut off point of -1.7 and 1.7, respectively, was then defined to select the stimuli. A significant difference was obtained between the scores of "aggressive" words ($M = 1.95$, $SD = .74$) and "peaceful" words ($M = -1.38$, $SD = .63$), $t_{(1,85)} = 13.91$, $p < .001$.

To obtain a pool of items with an equally extreme position on both sides of the scale, "peaceful" words and "aggressive" words were matched for rated extremity, i. e. for an "aggressive" verb with a value of 1.50, e. g., corresponding "peaceful" verb with a value of -1.50 was selected. Thus, 10 aggressive and 10 peaceful words resulted. The final material is shown in the appendix.

The stimuli for the target categories consisted of male and female Christian names and professions (see appendix).

Response categories

The target categories were defined by the concepts "male" and "female". The attribute categories were "rough" vs. "gentle" and consisted of the aggressive and peaceful words described above.

Participants

26 students enrolled in classes of Psychology, Computer Science, Geology and Education at Leipzig University participated in the experiment. 4 participants were community members. The subjects, 14 female and 16 male, were aged from 16 to 58 years ($M = 26.96$, $SD = 9.38$). Some of them were paid € 6 for their participation.

Design

The basis of the experiment was a 2 (IAT gender stereotype: male + rough vs. female + rough) x 2 (exercise: repetition of compatible and incompatible conditions) x 2 (counterbalancing factor: first vs. second part with replication of the IAT and order of compatible and incompatible conditions reversed) within-subjects factorial design. The between-subjects factors consisted of subject sex and the sequence of IAT conditions (male + rough first vs. female + rough first) (cf. Table 1).

Procedure

At the beginning the participants received a short introduction to the experiment described as speed of word processing. Detailed instructions for the task were given on the computer screen.

In the first form the IAT consists of five consecutive blocks (Greenwald, McGhee, & Schwartz, 1998a). In principle, five different blocks are used in our experiment as well. In the first block the subjects have to perform a target discrimination task, in this case to discriminate between male and female names by pressing a response button on their left side for male and on their right side for female names. The following task is an attribute discrimination task (left button for "rough", right button for "gentle" words). Next, a combined categorical decision is called for (left for male names and aggressive verbs, right for female names and peaceful words; compatible condition). In the next block the response assignment of the attribute categories is reversed (left button for "gentle", right button for

”rough” words). The last block consists of a reversed combined discrimination task (left for male names and peaceful verbs, right for female names and aggressive words; incompatible condition).

To examine the contribution of switch-costs (Mierke & Klauer, 2001) and exercise effects, a new variant of the IAT with seven blocks was used here. In this version of the IAT the compatible and the incompatible tasks were repeated. To analyse possible order effects, two groups of participants were formed. In the first part each group started either with the compatible or with the incompatible condition. In the second part of the experiment all seven stages from the first half were replicated, with the order of the compatible and the incompatible conditions reversed between groups (see Table 1).

TABLE 1: Design and Experimental Conditions (Seven-Stage Variant)

		Task Sequence 1		Task Sequence 2		
		blocks	left button	right button	left button	right button
Part 1	1		female	male	female	male
	2		gentle	rough	rough	gentle
	3		female / gentle	male / rough	female / rough	male / gentle
	4		female / gentle	male / rough	female / rough	male / gentle
	5		rough	gentle	gentle	rough
	6		female / rough	male / gentle	female / gentle	male / rough
	7		female / rough	male / gentle	female / gentle	male / rough
Part 2	8		female	male	female	male
	9		rough	gentle	gentle	rough
	10		female / rough	male / gentle	female / gentle	male / rough
	11		female / rough	male / gentle	female / gentle	male / rough
	12		gentle	rough	rough	gentle
	13		female / gentle	male / rough	female / rough	male / gentle
	14		female / gentle	male / rough	female / rough	male / gentle

Note: blocks 3, 4, 13, and 14 represent the compatible condition, blocks 6, 7, 10, and 11 form the incompatible condition.

Familiarity effects with respect to the IAT procedure should be captured by this replication. Subjects were randomly assigned to one of these two groups.

On the appearance of a stimulus in the center of the computer screen, the participants were asked to press the appropriate key as fast as possible. The order in which the words were presented was randomized for each subject. To reduce the number of classification errors and to increase the precision of the experiment, an error message (a red letter "X") appeared above the stimulus word after a wrong classification. The stimuli remained visible on the

screen until the subject responded correctly. During the whole experiment the category labels (i.e. *male vs. female* or *gentle vs. rough*) appeared in the upper left or right area of the screen.

At the end of the experiment, all participants filled out a questionnaire where they rated the material used in the experiment on a seven point "Likert - scale" from -3 to 3 with the endpoints labelled as "peaceful" and "aggressive". In addition, the participants filled out a German version of the "Modern Sexism Scale" (MSS, Eckes & Six - Materna, 1998).

After a short explanation of the true goal of the experiment and after answering their questions, the participants were dismissed.

Results

Data reduction. Erroneous classifications were not included in the analysis of reaction times. Altogether, 6.4% classification errors were made in the course of the experiment. These errors were evenly distributed over the experimental conditions. Reaction times linked to classification errors could therefore be eliminated without distorting the results. In accordance with previous suggestions (Rudman et al. 1999) reaction times less than 300 ms and over 3000 ms were excluded from the analysis (0.49% of the correct answers). Individual reaction times were log transformed and then averaged.

Implicit Association test. To test the hypotheses an ANOVA (2 x 2 x 2 x 2 x 2) was calculated with the within-subjects variables "gender stereotype" (male + aggressive vs. female + aggressive), "exercise" (repetition of the compatible and incompatible conditions) and "replication" (first vs. second test half with order of compatible and incompatible conditions reversed; counterbalancing factor). The between-subjects factors were "subject sex" and "sequence of IAT conditions" (male + aggressive first vs. female + aggressive). The α - level was set to .05. All calculations were done with the log-transformed reaction time data. Mean reaction times given refer to untransformed values in ms. Only the data in the relevant conditions (combined compatible and incompatible conditions) were analyzed.

Testing the hypothesis that aggressiveness is more strongly associated with men (compatible condition) than with women (incompatible condition), the main effect of the factor "gender stereotype" (IAT effect) becomes significant ($F_{(1,26)} = 19.86$, $p < .001$, effect size $d = 2.07$). The reaction time means in the compatible and the incompatible condition were 693 ms (SD = 20.2) and 794 ms (SD = 24.9), respectively, resulting in an IAT effect of 101 ms. All means are shown in Figure 1 (Appendix). In agreement with the hypothesis no

interaction effect for gender stereotype with subject sex was obtained ($F_{(1,26)} = 1.77, p = .195$). This means that the magnitude of the IAT effect was the same for male and female subjects. Main effects are observed for the within-subject factors "exercise" ($F_{(1,26)} = 24.58, p < .001$, effect size $d = 1.14$, $M = 758$ ms ($SD = 22.4$) when a combined condition is taken for the first time, and $M = 729$ ms ($SD = 21.2$) when taken for the second time), "replication" ($F_{(1,26)} = 45.96, p < .001$, effect size $d = 1.87$, $M = 768$ ms ($SD = 23.7$) for first test half, and $M = 719$ ms ($SD = 19.9$) for second test half), and for the between-subjects factor "subject sex" ($F_{(1,26)} = 14.87, p = .001$, effect size $d = 5.5$; $M = 827$ ms, $SD = 31.4$ for male subjects, and $M = 660$ ms, $SD = 29.4$ for female subjects). That means, in agreement with hypotheses no. 3a and 3b, training or exercise effects do exist for the repetition of combined conditions (the factor exercise) and for the replication of the whole IAT (first vs. second test half, the factor replication). The latter effect consists of an improvement from the first to the second half, which indicates a greater familiarity with the procedure of the IAT when taking the test a second time. The former effect, based on the replication of the combined conditions, seems to reflect an increase in simple motoric convenience, i.e. a motoric exercise effect. The group factor "sequence of IAT conditions" does not have a significant effect ($F_{(1,26)} = 2.02, p = .170$). Therefore, it does not matter, which of the combined conditions, the compatible or the incompatible one, is presented first. A significant interaction appeared for the IAT effect with the replication factor (first vs. second test half, $F_{(1,26)} = 54.93, p < .001$), with the IAT effect diminishing from the first ($M = 123$ ms) to the second test half ($M = 80$ ms), and for the replication factor with the exercise factor (repeating of the combined conditions, $F_{(1,26)} = 15.45, p = .001$). However, contrary to hypothesis no. 4 (task switching hypothesis) no interaction effect was obtained between IAT effect and the group factor sequence of combined conditions ($F < 1.0$), nor was there a three-factor interaction between task sequence, exercise (replication), and gender stereotype ($F < 1.0$). Thus, there is no evidence for an influence of the task-switching factor on the IAT effect. Testing the interaction of task sequence (compatible vs. incompatible condition first) with the exercise factor, a non-significant F-statistic resulted ($F < 1.0$). The simple interaction effect of task sequence with exercise was marginally significant, however, for the first test half only ($F_{(1,26)} = 3.83, p = .061$), with the exercise effect being much larger when the compatible condition preceded the incompatible one compared to the reverse sequence.

The results of the ANOVA pertaining to the hypotheses are summarized in Table 2.

TABLE 2: ANOVA of Experiment 1

	F	df	p
Effect			
Exercise	24.58	1,26	.000
Replication	45.96	1,26	.000
Gender Stereotype (IAT Effect)	19.86	1,26	.000
Subject Sex	44.84	1,26	.000
Block Sequence	2.02	1,26	n. s.
Gender Stereotype * Subject sex	1.77	1,26	n. s.
Gender Stereotype * Replication	54.93	1,26	.000
Exercise * Replication	15.45	1,26	.001
Exercise * Block Sequence	3.83	1,26	.061
Gender Stereotype * Block Sequence	< 1	1,26	n. s.
Gender Stereotype * Block Sequence * Replication	< 1	1,26	n. s.

Precision of Measures. To test the reliability of the measure, the retest – reliability and the internal consistency (Cronbach α) were calculated. In agreement with similar results (e.g. Greenwald, 2001) extremely high internal consistencies ($\alpha = .98$) were obtained. The retest – reliability for the evaluatively compatible condition was $r = .959$ ($p < .001$), for the evaluatively incompatible condition $r = .909$ ($p < .001$), and for the IAT – effect $r = .723$ ($p < .001$).

Explicit measures. At the end of the session the participants rated the experimental material used (attribute stimuli) on a seven point "Likert - scale" from -3 to 3 with the endpoints labelled as "peaceful" and "aggressive", as had already done the participants in the preliminary study. Again, a large difference between "aggressive" words and "peaceful" words was obtained ($t_{(1,29)} = 32.1$, $p < .001$). With respect to the evaluation of the stimuli and the scores of the "modern sexism scale", no statistically significant differences for men and women were found.

Correlation between measures. In agreement with the literature (e.g. Greenwald et al. 1998b), no significant correlations between the IAT – effect and explicit measures were observed. However, all coefficients were in the expected direction (e.g. IAT – effect and rating of "aggressive"- verbs $r = .317$, n. s.). The only significant correlation obtained occurred within the explicit measures between the rating of "aggressive"- and "peaceful" verbs ($r = -.466$, $p = .012$), but neither for the "Modern Sexism Scale" (MSS) nor for the IAT – effect any significant correlation resulted. All correlations are shown in Table 3.

TABLE 3: Correlation Among Measures (experiment 1)

	AGGR	PEACE	MSS	IAT – effect
AGGR	1.000 .	-.466* (.012)	-.219 (.262)	.317 (.100)
PEACE		1.000 .	.068 (.722)	-.200 (.289)
MSS			1.000 .	-.083 (.664)
IAT – effect				1.000 .

* Correlation is significant (5 %) / N = 30

in brackets: probability

Notes: AGGR = Rating of the aggressive verbs

PEACE = Rating of the peaceful verbs

MSS = "Modern Sexism Scale"

Discussion

This study successfully demonstrates the usefulness of the IAT for the assessment of implicit attitudes to aggression. Most of the hypotheses set up in advance could be confirmed. First, the strong effect of the factor "gender stereotype" showed that the implicit association of men with aggression is stronger than the association of women with aggression. Second, this effect was of the same magnitude for male and female subjects. That means, according to our data, there was no evidence for ingroup-outgroup differences.

Third, the differences between the first and the second presentation of the test show that the IAT effect can not be explained by implicit attitudes alone. Moreover, a larger exercise effect was observed in the first test half relative to the second half when repeating the compatible and the incompatible conditions. The IAT effect (stereotype effect) was still present in the second presentation of the test, when it was replicated. If the IAT had measured implicit attitudes alone and implicit attitudes are "outside conscious awareness and control" (Cunningham et al., 2001; Banaji, 2001), then a replication of the test procedure should not lead to a decrease in the effect size. When such a decrease in effect size is obtained, it could be explained in two ways. Either it could be assumed that the implicit attitude has changed from first to second presentation of the IAT, or one can assume that other factors like unfamiliarity of the test procedure are responsible for the larger effect in the first presentation. The first explanation doesn't seem very plausible, because implicit attitudes are considered as relatively "time stable constructs" (see Greenwald & Banaji, 1995) and implicit measures

may be taken as stable indicators of the corresponding attitudes (Cunningham et al., 2001). Therefore, the assumption of procedural factors leading to a lower effect size in second presentation of the IAT remains as an explanation for our results. This decrease in effect size can be traced back either to switch-cost effects (Mierke & Klauer 2001) or simple exercise effects. In support of the latter interpretation as exercise effects one may refer to processing mechanisms like the cognitive capacity needed for performing the decision task, which will decrease with sufficient exercise. It should be examined in further experiments which of the two factors, switch costs or exercise, has a stronger influence on the IAT effect.

Fourth, there was no significant effect for the presentation order of the combined conditions, so that the hypothesis the size of the IAT effect would depend on the presentation order of the conditions has to be rejected. Thus, it does not matter, whether the participants start with the compatible condition or with incompatible condition.

The results further show that the seven stage variant of the IAT is suitable to detect exercise effects if they are present. It seems that the second presentation of the combined conditions of the IAT would lead to a more appropriate measure of the implicit attitudes. By replicating the whole IAT procedure for a second time a smaller IAT effect was obtained, which indicates that the training effects are reduced in the second test presentation but that nevertheless a substantial stereotype effect still remains.

Study 2

Study 2 was designed as a replication of the results of study 1, in particular to confirm the exercise effects obtained and to establish the usefulness of the seven-stage variant of the IAT. That means, the compatible and the incompatible conditions are repeated once for every subject. A response window version of the IAT was used here. By this procedure the participants are forced to respond within a limited period of time, the response windows, in our case within 1500 ms. During this period the stimulus is visible all the time, and the total duration of the stimulus presentation is indicated by a red bar increasing in length which appears on the upper screen edge. Reactions above 1500ms were counted as errors. After the stimuli had disappeared from the screen, answers could still be given, but these data were not included in the analysis. In addition, implicit self-esteem instead of aggression was chosen as the attitude domain. According to Greenwald and Farnham (2000), the IAT is also appropriate to investigate the attitude towards the self. It can be assumed that the self-category is more strongly associated with positive attributes, whereas negative attributes should be more

strongly associated with the category of "other" (non self). As no order effects were obtained in Experiment 1, the sequence of IAT conditions was not varied.

The following hypotheses will be investigated:

1. The implicit association of self is greater with positive than with negative attributes (IAT effect, compatible vs. incompatible response assignment).
2. No gender differences do exist for the IAT effect.
3. Exercise is expected to reduce the IAT effect.

Method

Material

The stimuli belonging to the target category "self" were self-related pronouns (I, mine, my, self) and individual data characteristic for a particular subject (her or his month, year, city and country of birth, Christian name, family name, age, sex), forming a total of 13 stimuli. The complementary target category of "other" was defined by other-related words (they, their, your, others) and the corresponding characteristics differing from the personal data of a subject (see in the appendix). The attribute stimuli consisted of words with a positive (bright, proud, nice, success) or a negative connotation (dull, ugly, useless, loss). Before the experiment started, the participants entered their individual data into the computer. The computer then generated a personalized data file with the target stimuli presented to them.

Participants

81 participants (65 female and 16 male) aged 19 to 47 years ($M = 23.13$, $SD = 5.86$) took part in the experiment. The subjects, all psychology undergraduate students at Leipzig University, were enrolled in a course on experimental psychology.

Design

The basis of the experiment was a $2 \times 2 \times 2$ design. Within-subjects factors were exercise (repetition of the compatible and the incompatible conditions), and self-evaluation (IAT effect self + positive vs. self + negative). Sex of the participants was the between-subjects factor.

Procedure

The participants got a short introduction to the test procedure and to the tasks to be accomplished on the computer screen at the start of the experiment and at the beginning of every block. During the whole session the experimenter was available to answer questions.

Except from the material and the number of blocks to be completed, the procedure was the same as in study one. That means, the IAT consisted of 7 blocks with the compatible and the incompatible conditions each repeated once. Respectively, the stimuli had to be classified by the participants as "self" vs. "others" (target categories) and "gentle" vs. "rough" (attribute categories). In order for the participants to pay attention to a correct decision and not just to the speed, an error message (a red letter "X") appeared above the target after a wrong classification. The target remained visible on the screen until the subject responded correctly.

The order of the stimuli was randomized for every participant. Each stimulus within a category (attributes and targets) was presented once. In the combined response conditions (blocks 3, 4, 6, and 7), every target stimulus and every attribute stimulus appeared once in random order. Reaction times and error rates were analysed for the combined response conditions, compatible and the incompatible blocks, only.

After a short explanation of the goal of the experiment and answering their questions the participants were dismissed.

Results

Data reduction. In the critical blocks (combined response conditions) 12.2% of the classifications made were counted as errors on the average. These errors include correct decisions when the response deadline was exceeded (3.84% of all relevant cases). The reaction time data corresponding to classification errors were excluded from the analysis. The reaction time data were log-transformed before an analysis of variance was computed. For illustrative purposes, the untransformed reaction time means are given in the text.

Implicit Association test. In order to test the hypotheses, an ANOVA (2 x 2 x 2) with the within-subjects factor "exercise" (first vs. second presentation of combined response conditions) and self-evaluation (IAT effect self + positive vs. self + negative), and the between-subject factor "subject sex" was calculated.

In accordance with our hypothesis, the main effect of the self-evaluation factor ("IAT effect") is significant ($F_{(1,79)} = 258.1, p < .001$, effect size $d = 14.63$): positive stimuli are assigned to the category "self" ($M = 689$ ms, $SD = 11.4$) more quickly than to the category

”other” ($M = 876$ ms, $SD = 14.1$). All means are shown in Figure 1 (Appendix). As expected, this effect is not qualified by subject sex ($F_{(1,79)} = 1.29$, $p > .10$). A main effect for subject sex was not observed either ($F = < 1.0$). As is assumed by hypothesis three, a main effect due to exercise can be shown for the repetition of the combined response conditions ($F_{(1,79)} = 54.70$, $p < .001$, effect size $d = 3.61$). Upon the first presentation of a condition the mean reaction time was $M = 804$ ms ($SD = 11.7$), which reduced to $M = 761$ ms ($SD = 11.8$) on the second presentation. This time, however, a significant interaction effect between the self-related IAT effect and the exercise factor was obtained ($F_{(1,79)} = 8.80$, $p = .004$, effect size $d = 3.17$). The IAT effect slightly increased on the repetition of the IAT conditions from 174 ms in the first presentation to 199 ms in the second presentation. In addition, a second-order interaction between subject sex, exercise, and self-evaluation was observed ($F_{(1,79)} = 8.76$, $p = .004$, effect size $d = 11.9$), resulting from the fact that the IAT effect significantly increased during the repetition for the male subjects (from 146 ms to 204 ms) but remained practically constant for the female subjects (from 201 ms to 195 ms). As we had no hypotheses with respect to these effects, we will not discuss the latter results any further. All other effects of the design were insignificant (see Table 4).

TABLE 4: ANOVA of Experiment 2

	F	df	P
Effect			
Repetition	54.7	1,79	.000
Self Evaluation	258.1	1,79	.000
Subject Sex	< 1	1,79	n. s.
Self Evaluation * Repetition	8.8	1,79	.004
Self Evaluation * Subject Sex	< 1	1,79	n. s.
repetition * Subject Sex	< 1	1,79	n. s.
Self Evaluation * Repetition * Subject Sex	8.8	1,79	.004

Quality of measurement. To test the quality of the measurement, the retest – reliability and the internal consistency (Cronbach α) were calculated. A high internal consistency ($\alpha = .86$) was obtained for both conditions. The retest – reliability for the evaluative compatible condition is $r = .845$ ($p < .001$) and for the evaluative incompatible condition $r = .757$ ($p < .001$).

Discussion

All hypotheses set up in advance could be confirmed. A strong main effect of the self-evaluation factor showed that the implicit association of the self is more positive than the association of the category "others" with positive attributes. A main effect for subject sex was not observed, nor was the IAT effect qualified by subject sex. This is in agreement with our second hypothesis, which assumes that women assign positive attributes to self and negative attributes to others in the same way as men do.

The significant main effect when repeating the combined response conditions indicate a global decrease in reaction time due to an increase in motor exercise or familiarity with the test procedure. The interaction effect between the repetition factor and the IAT conditions which was observed here, however, uniquely results to the male subjects who showed an increase in the IAT effect with greater exercise, whereas the IAT effect of the female subjects remained unchanged independent of exercise. Once more the results allow the conclusion that the seven stage variant of the IAT suggested here is suitable to detect exercise effects if they are present. In future research we therefore recommend the seven-stage variant of the IAT in order to test for exercise effects.

General discussion

In two experiments we wanted to investigate the influences of exercise effects in the IAT and to find out if they are independent of the association between target categories and associated attributes. The two studies successfully confirmed the appropriateness of the IAT for the assessment of implicit self-value and aggression. In the second study stronger reaction time differences (effect size $d = 2.29$) were obtained compared to those in study one (effect size $d = 1.35$). This difference can be traced back to the differences in the arousing quality of the material. We may assume that measuring the self-evaluation or the self-image of an individual leads to a stronger activation than the measurement of gender stereotypes even if aggression is concerned. In comparison to the results of other authors, e. g. Greenwald et al. (1998a) who report effect sizes of $d = 1.21$ on the average, we obtained larger effect sizes in the present experiments.

The estimation of the reliability of our IAT measures show an internal consistency (Cronbach α) of $\alpha = .86$ in experiment 2 and an $\alpha = .98$ in experiment 1. Both values are much higher than those reported by Cunningham et al. (2001) who obtained an internal consistency of $\alpha = .78$ which they consider to be in an acceptable range. In addition to

internal consistency, the measure of a construct like implicit attitudes should be stable over time. Cunningham et al. (2001) report a very low value test-retest – reliability of $r = .27$ for the IAT. They explain the low stability in their experiments by referring to a contamination with measurement errors. Low stability over time may be a severe problem for implicit attitude measures indeed. In this context the test-retest – reliability obtained here in experiment 2 is sufficiently high ($r = .845$ for part 1, $r = .757$ for part 2) and is even higher in experiment 1 ($r = .959$, and $r = .909$).

Although the effect sizes strongly differ between the two experiments, strong effects due to a repetition of IAT conditions are evident in both studies, too. As already mentioned, an attitude change during the course of the experiment would seem to be improbable. The high values of the test-retest – reliability and the internal consistency (Cronbach α) of the data in our experiments confirm this assumption. If the difference in reaction time between the evaluatively compatible and the evaluatively incompatible conditions is due to the underlying implicit attitude, and if this attitude does not change during the experiment, other factors for the variation of the IAT effect with different procedural conditions (e.g. sequence of conditions, exercise, test replication) must be considered. Mierke and Klauer (2001), e. g., interpret this variation in reaction time at least partly of so called switch-costs resulting from a change between compatible and incompatible conditions. If the account in terms of switch-costs were correct, the magnitude of the effect should differ depending on the particular sequence of the compatible and the incompatible conditions. When changing from a less complex condition (i.e. the evaluatively compatible one) to a more complex condition (i.e. the evaluatively incompatible condition), more switch costs would arise compared to a change from a more complex to a less complex condition. Therefore, the IAT effect should be different depending on which one of the two conditions appears first. However, such a sequence effect was not observed in our first experiment. Instead of this account of the IAT effect given by Mierke and Klauer (2001) we would favour a different kind of explanation. If the incompatible condition precedes the compatible one, more cognitive resources will have to be released to accomplish this relatively complex task. Thus, a change from the incompatible to the compatible task condition would be easier to accomplish than a change from the compatible to the incompatible task. This sequence of conditions should therefore lead to greater exercise effects than in case of the compatible condition coming first. Up to this point of the argumentation, the results of Greenwald et al. (1998), and Mierke and Klauer (2001), could still be interpreted in terms of switch-costs. However, it should be clear that

switch-costs can only arise, when the mapping of the response categories changes from one condition to the other one, but not when the same condition is repeated. If the reaction time changes in the latter case, then this effect cannot be explained by switch-costs any more. According to our interpretation then, just exercise effects and not task-switch costs would be responsible for the variation of the IAT effect. However, according to the interpretation in terms of an exercise effect, an increase or decrease of the IAT effect may or may not appear, depending on the kind of material used or the kind of attitude domain investigated. Switch costs, on the other hand, will always be involved in this kind of decision task, but will be independent of the kind of material or the attitude domain.

As shown in our experiments, the IAT was appropriate for measuring implicit attitudes towards aggression as well as assessing the implicit self concept. In any case, the assumption that the IAT effect represents implicit attitudes alone seems to be problematic. However, in accordance with the view of Greenwald and Nosek (2001), the IAT can be recommend for practical purposes, if some of the procedural factors having an influence on the magnitude of the effect are taken into account.

The presented experiments also made evident that we still need to know more about the cognitive processes underlying the IAT effect in order to understand what the procedure really measures. We were able to show that exercise had a large impact on the magnitude of the IAT effect, which is independent of the strength of the association between target categories and the associated attributes.

References

- Banse, R., Seise J. & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, 48, 145-160
- Cunningham, W.A., Preacher, K.J. & Banaji, M.R. (2001). Implicit attitude measures: consistency, stability, and convergent validity. *Psychological Science*, 2, 163-170
- De Houwer, J. (2001). A Structural and Process Analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, 6, 443-451
- Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C. & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238
- 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 a). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A. & Rosier, M. (1998 b). Prologue of a unified theory of attitudes, stereotypes, and self-concept. In J.P. Forgas (Ed.), *Feeling and Thinking: The Role of Affect in Social Cognition and Behavior*. New York: Cambridge University Press
- Greenwald, A. G. & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022-1038
- Greenwald, A.G. & Nosek, B.A. (2001). Health of the Implicit Association Test at Age 3. *Zeitschrift für Experimentelle Psychologie*, 48, 85 - 93
- McConnell, A. & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435-442
- Mierke, J. & Klauer, K. C. (2001). Implicit Association Measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie*, 48, 107-122
- Rothermund, K. & Wentura D. (2001). Figure ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie*, 48, 94-106

Rudman, L.A., Greenwald, A.G., Mellott, D. S. & Schwartz, J. L. (1999). Measuring the automatic components of prejudice: flexibility and generality of the Implicit Association Test. *Social Cognition*, 4, 437-465.

Correspondence to this article may be addressed to:

Ronny Werner, Institute for Cognitive Psychology, Social Psychology, University of

Leipzig, Seeburgstr. 14/20, 04103 Leipzig, Germany

e-mail: post@ronnywerner.de

Appendix

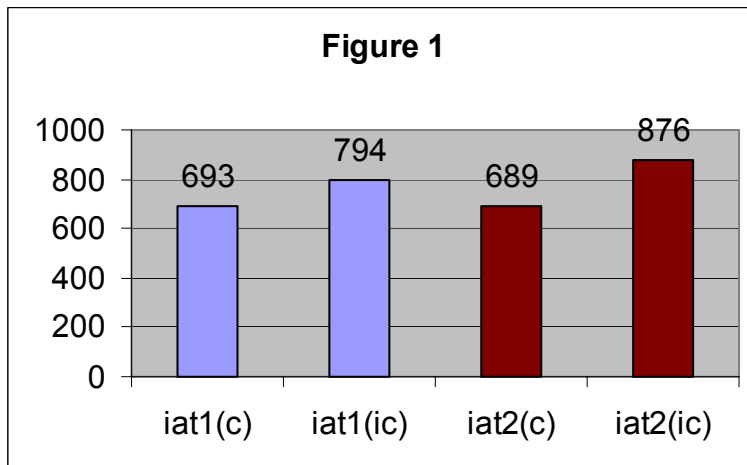
"peaceful" words			"aggressive" words		
Item	M	SD	Item	M	SD
Adaptive	- 2,16	.14	Exploding	+ 2,17	.15
Giving	- 1,85	.13	Offending	+ 1,73	.10
Helping	- 2,33	.13	Stepping	+ 2,26	.11
Caressing	- 2,40	.11	Beating	+ 2,33	.11
Relaxing	- 2,09	.14	Threatening	+ 2,07	.09
Sleeping	- 2,12	.18	Discriminating	+ 2,12	.10
Mediating	- 2,00	.12	Shouting	+ 2,00	.11
Trusting	- 1,95	.18	Raging	+ 1,92	.12
Allocating	- 2,35	.13	Hating	+ 2,29	.15
Resting	- 2,26	.16	Tormenting	+ 2,26	.15

Material of Experiment 1: Attribute Stimuli

		Professions		Christian names	
		male	female	male	female
1	Doctor		Woman doctor	Lutz	Ulrike
2	Orderly		Nurse	Jörg	Astrid
3	Secretary		Woman Secretary	Holger	Annett
4	Teacher		Woman Teacher	Jens	Antje
5	Seller		Woman Seller	Klaus	Maria

Material for Experiment 1: Target Stimuli

IAT - means for Experiment 1 & 2



c = compatible condition
ic = incompatible condition