The
British
Psychological
Society

www.bpsjournals.co.uk

The implicit association test outperforms the extrinsic affective Simon task as an implicit measure of inter-individual differences in attitudes

Jan De Houwer* and Els De Bruycker
Ghent University, Ghent, Belgium

We used both the Extrinsic Affective Simon Task (EAST) and the Implicit Association Test (IAT) as implicit measures of inter-individual differences in attitudes towards political parties (Experiment 1), food items (Experiment 2) and homosexuality (Experiment 3). IAT but not EAST scores were related in a meaningful manner to self-report measures of the corresponding attitudes (Experiments 1–3) and self-reported behaviour (Experiments 2 and 3). Whereas split-half reliability of the IAT scores was satisfactory, EAST scores overall had a low split-half reliability. The present results suggest that the EAST as introduced by De Houwer (2003b, *Experimental Psychology*) does not offer a good alternative for the IAT as an implicit measure of inter-individual differences in attitudes.

During the past decade, implicit measures of attitudes have become popular in various subdisciplines of psychology (see Fazio & Olson, 2003, for a review). Implicit measures of attitudes can be defined as measures that capture attitudes in an automatic manner (De Houwer & Moors, 2007). These measures appeal to many researchers because they have potential advantages compared with more traditional explicit measures. The most often cited advantages are that implicit measures might be less biased by deliberate attempts to conceal the attitude and that they might even reflect attitudes of which the respondent is not aware (e.g. Dovidio & Fazio, 1992; Greenwald & Banaji, 1995).

Although several implicit measures have been proposed over the past 10 years (see De Houwer, 2003a; Fazio & Olson, 2003, for reviews), the Implicit Association Test (IAT) that was introduced by Greenwald, McGhee, and Schwartz (1998) is beyond any doubt the most popular implicit measure. In IAT studies, participants see stimuli that are presented on a screen one by one. The stimuli belong to one of four categories and need to be categorized by pressing one of two keys. Results typically show that responses are faster when associated categories are assigned to the same key than when they are assigned to different keys. Assuming that attitudes are represented in memory as associations between target

*Correspondence should be addressed to Jan De Houwer, Department of Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium (e-mail: Jan.DeHouwer@UGent.be).

concepts and the concepts 'positive' and 'negative', the IAT can be used to measure attitudes; for instance, take the racial IAT that was developed to measure attitudes towards Black and White people. On each trial, a positive word, a negative word, a name typical of a Black person (e.g. Leroy; Black names) or a name typical of a White person (e.g. Hank; White names) is presented on the screen. In the Black-positive task, participants are asked to press one key for both positive words and Black names and the second key for negative words and White names. Participants also complete a White-positive task in which they press one key for positive words and White names and the second key for negative words and Black names. To the extent that a participant has a more negative attitude towards Black people than towards White people (i.e. relative to the concept 'White person', the concept 'Black person' is associated more with negative than with positive valence), he or she should find the Black-positive task more difficult than the White-positive task. Hence, the difference in performance between the two tasks (i.e. the IAT effect or score) can be taken as an index of racial attitudes.

Many studies have now confirmed that IAT scores can provide a valid and reliable measure of inter-individual differences in attitudes. For instance, studies have shown that inter-individual differences in IAT scores are related in a meaningful manner to group membership (e.g. Greenwald *et al.*, 1998), attitude-related behaviour (see Poehlman, Uhlmann, Greenwald, & Banaji, 2005, for a review) and, under certain conditions, self-reported attitudes (e.g. Nosek, 2005; see Nosek, Greenwald, & Banaji, 2007, for an excellent review of research on the IAT). However at the same time, it is clear that the IAT has its problems and limitations (see De Houwer, 2002; Nosek *et al.*, 2007, for a discussion). For instance, the IAT can reveal only relative attitudes. Consider the racial IAT that we described above. It can, at best, reveal the attitude towards Blacks relative to the attitude towards Whites. It cannot reveal whether the attitude towards Blacks (or Whites) is in itself positive or negative (see Wigboldus, Holland, & van Knippenberg, 2005, for a possible way to overcome this limitation). Another potential problem is that IAT effects can be influenced by factors other than personal attitudes (e.g. Olson & Fazio, 2004; Rothermund & Wentura, 2004). These problems have stimulated researchers to look for alternative measures that could be used instead of or in addition to the IAT.

One of these alternative measures is the Extrinsic Affective Simon Task (EAST) that was introduced at our laboratory by De Houwer (2003b; see De Houwer & Eelen, 1998, for a precursor). In a typical EAST study, words that are coloured white, blue or green are presented one by one on a screen. As in the IAT, participants are asked to respond to these words by pressing one of two keys. When the word is white, participants are asked to respond on the basis of valence by pressing one key for positive white words and another key for negative white words. As a result of these trials, one key becomes associated with positive valence and the other key with negative valence. When the word is coloured, participants need to select one of those same responses, but this time based on the colour of the word. De Houwer (2003b) showed that responses to positive coloured words were faster when participants pressed the positive key (i.e. the key that was also assigned to positive white words) than when they pressed the negative key (i.e. the key that was also assigned to negative white words) whereas the reverse was true for negative coloured words. In principle, the EAST can be used to measure attitudes. For instance, in order to measure self-esteem (i.e. attitudes towards the concept 'self'), one can present coloured words that are related to the concept 'self' (e.g. the first name of the participant or words such as I or MYSELF). If a participant needs less time to give positive responses to those words than to give negative responses, this would indicate that the participant has positive self-esteem (see De Houwer, 2003b).

De Houwer (2003b) argued that the EAST might overcome some of the problems of the IAT. First, the EAST, in principle, allows one to measure single or multiple attitudes in a non-relative manner. For each target concept, one can present representative stimuli and compare the time to give positive responses with the time to give negative responses to those stimuli. Second, because the EAST measure is based on a comparison of trials within a single task rather than on a comparison of performance in two different tasks (as is the case with the IAT), EAST measures might be less sensitive to non-attitudinal factors that participants can use to recode the different IAT tasks (see De Houwer, 2003b). However, De Houwer (2003b) did not present data that supported the claim that the EAST can provide a valid and reliable measure of inter-individual differences in attitudes. He did demonstrate that mean EAST effects as averaged across participants were in line with the normative valence of the items that were presented. However, this does not imply that inter-individual differences in EAST scores provide a valid and reliable indication of inter-individual differences in attitudes.

There are some indications that the EAST, as introduced by De Houwer (2003b), might not be particularly suited for measuring inter-individual differences. First, De Houwer himself noted that the split-half reliability of inter-individual differences in EAST scores tended to be low. Second, studies by Teige, Schnabel, Banse, and Asendorpf (2004) and by Schmukle and Egloff (2006) showed that the reliability of EAST scores was far lower than the reliability of IAT scores and that the IAT was a far better predictor of criterion variables than the EAST. However, in these studies, the EAST was used to measure aspects of personality (e.g. associations between the concept 'self' and the concept 'anxiety') rather than attitudes, as was the original purpose of the EAST. It thus remains to be seen whether the EAST, as introduced by De Houwer, can provide a good measure of inter-individual differences of attitudes and how the EAST compares with the IAT in this respect. We therefore conducted a number of studies in which we used both the EAST and the IAT to measure inter-individual differences in attitudes towards a variety of attitude objects.

In our first experiment, we measured attitudes towards two political parties. Based on a pilot study, we selected parties for which we could expect large inter-individual differences in attitudes among our participants. This is important because, in order to measure inter-individual differences, there of course need to be inter-individual differences in the first place. A second reason for selecting these attitude objects is that one can expect strong correlations between implicit and explicit (i.e. self-reported) attitudes towards political parties (e.g. Nosek, 2005). This allowed us to test the validity of our EAST and IAT measures by assessing the strength of the correlation between the EAST and IAT measures on the one hand and self-reported attitudes on the other hand.

The second experiment was analogous to the first one except that we now measured attitudes towards the concepts 'beer' and 'sprouts' (which is a cabbage-like, small green vegetable). Again, a pilot study indicated that there was a large variation in attitudes towards these attitude objects. Since there are no obvious reasons why participants would not reveal their true attitude towards these objects on self-report measures and because previous research showed fairly high implicit–explicit correlations for food items (e.g. Nosek, 2005), we again used self-reported attitudes as the criterion to assess the validity of our implicit measures. In addition, we also examined the relation between IAT and EAST scores on the one hand and self-reported frequency of consumption of beer and sprouts on the other hand.

In the third experiment, we adopted a known-group approach. Half of the participants said they were heterosexual whereas the other participants indicated that they were homosexual. The validity of the implicit measures was assessed by examining

the extent to which the hetero- and homosexual participants had different attitudes towards homo- and heterosexuality as indexed by our measures.

EXPERIMENT 1

The target objects in this experiment were the Flemish political parties VLD and CD&V. VLD is the party of the Flemish liberals. At the time of the experiment, VLD was part of a coalition government with the Socialist and Green party at the federal Belgian level and the regional Flemish level. CD&V is the party of the Flemish Christian Democrats. At the time of the experiment, they were the main opposition party at both the federal and regional level. The experiment was conducted 3 months before federal elections were held. All Belgian citizens older than 18 years were required to take part in these elections. We chose VLD and CD&V because they were well-known political parties that have dominated recent Belgian politics. In addition, a pilot study indicated that psychology undergraduates at Ghent University varied strongly in their self-reported attitudes towards these parties.

Method

Participants

Forty-eight undergraduate psychology students at Ghent University were paid €5 for their participation in this experiment. They were between 18 and 23 years old and thus eligible to vote. Participants were recruited using flyers that were distributed near classrooms. They took part in groups of 4–24 persons. As in all other experiments, participants were native Dutch speakers.

Measures and procedure

Before filling in an informed consent form, participants were told that the experiment consisted of two reaction time tasks on a personal computer and a short questionnaire. Half of the participants first filled out the questionnaire; the other half received the questionnaire after the two reaction time tasks were performed. Also, for half of the participants, the IAT was presented as the first implicit measure, whereas for the other participants, the EAST was the first implicit measure. The experiment was conducted in a dimly lit classroom containing 25 personal computers with Pentium processors and 15-inch screens. Participants were placed in front of one of the computers. All reaction time tasks in this and the other experiments were implemented using Inquisit 1.33 software. Questionnaires were printed on paper.

Self-report questionnaire

All participants filled in a short questionnaire. First, they rated on a seven-point scale (from −3 to 3) how much they liked VLD and CD&V. Afterwards, they were asked whether they voted for one of these two parties at the previous elections (past voting behaviour). Participants could chose between the responses 'yes', 'no', 'I can't remember' and 'I was not yet allowed to vote at the previous elections'. To assess voting intentions, for each party we asked whether they intended to vote for that party during the next elections. The possible responses were 'no', 'probably not', 'haven't thought about it', 'probably yes' and 'yes'.

Implicit association test

Target stimuli in the IAT were words, logos and colours referring to VLD and CD&V. As attribute stimuli, we used positive and negative words and pictures (see the Appendix for a description of all stimuli). Stimuli were presented on the computer screen on a white background. As in all other tasks and experiments, words were written in Arial Black font with font size 36. Attribute pictures were black and white. Target pictures were coloured. Participants could respond by pressing key q (left) or p (right) on a standard QWERTY keyboard.

Instructions at the start of the IAT informed participants that words and images would appear one by one on the computer screen in front of them. They were asked to press the left key (q) or the right key (p) depending on the category of the word or image. The four categories were VLD, CD&V, positive and negative. The key assignments would vary across different phases but the category labels printed at the left and right upper corner of the screen would indicate the correct key assignment during a particular phase. Participants were instructed to respond as quickly as possible without making too many errors and were informed that the task would take about 10 minutes.

The IAT started with an attribute practice phase of 24 trials during which each of the four positive and four negative stimuli was presented three times. This was followed by a target practice phase of 24 trials in which each target stimulus was presented three times. These two tasks were then combined in two test blocks of 48 trials each. During each test block, each target and attribute stimulus was presented three times. In a fourth phase, participants received a second target practice block of 24 trials, but now with reversed key assignments for the target categories. Finally, the two test blocks of Phase 3 were repeated but now with the reversed key assignments for target categories. The order of the trials was determined randomly for each block and participant separately. On each trial, the stimulus was presented in the centre of the screen until a response was registered. If the response was incorrect, a red X appeared in the middle of the screen for 400 milliseconds. The next trial started 400 milliseconds after the correct response was registered or the red X disappeared.

Before each block, instructions were given about the upcoming task. Half of the participants pressed the left key for positive words and the right key for negative words in all phases of the task. The other half received the opposite key assignment for the attribute categories. Orthogonal to this counterbalancing manipulation, half of the participants pressed the left key for CD&V and the right key for VLD during Phases 2 and 3. The other half received the opposite key assignment for the target categories during Phases 2 and 3. As a result, half of the participants first performed the task in which VLD and positive stimuli were assigned to one key and CD&V and negative stimuli were assigned to the second key (VLD-positive task) whereas the other participants first pressed one key for VLD and negative stimuli and the other key for CD&V and positive stimuli (CD&V-positive task).

Extrinsic Affective Simon Task

Target items were the letter strings 'CD&V' and 'VLD' printed in pink (RGB values of 250, 120, 210) and purple (RGB values of 210, 120, 250). We chose these letter strings because they most unequivocally represent the political parties. If we had also presented other stimuli, such as the colour associated with each party or the names of prominent members of each party, there was a risk that these stimuli would activate responses not based on the attitude towards the respective party but

based on other attitudes, such as attitudes towards the colours or persons as such. This risk is higher in the EAST than in the IAT because, unlike what is the case in the IAT, in the EAST, participants are not forced to encode the stimuli in terms of the target categories (see De Houwer, 2001, in press, for a discussion). Whereas De Houwer (2003b) used blue and green as colours for the coloured words, we chose pink and purple because the former but not the latter are colours associated with Belgian political parties. Attribute items were positive and negative words (see the Appendix) that were always presented in white upper case letters. All stimuli appeared on the computer screen on a black background. Again, participants responded by pressing either the q key or the p key on a QWERTY keyboard.

Participants were told that words would appear one by one on the computer screen and that their task was to press the negative or the positive key depending on the meaning or the colour of the words. For white words, they were asked to press the left key or the right key depending on valence. For coloured words, they were asked to press the left or right key depending on colour. Half of the participants pressed left for positive words and right for negative words whereas the reverse was true for the other participants. Orthogonal to this manipulation, half of the participants pressed the left key for pink words and the right key for purple words whereas the others pressed left for purple words and right for pink words. For half of the participants, the response mapping for the attribute stimuli in the EAST was the opposite to the mapping of the attribute stimuli in the IAT. The other half received identical attribute response mappings in both tasks. All participants were instructed to respond as quickly as possible without making too many errors and were informed that the task would take about 15 minutes.

The EAST started with an attribute practice phase that consisted of 24 trials during which each attribute word was presented three times. This was followed by a target practice block of 24 trials during which the stimuli VLD and CD&V were each presented six times in purple and six times in pink. Next, there were four mixed test blocks consisting of 48 trials. In each test block, the eight attribute words were presented twice in white and the two target stimuli were presented eight times in pink and eight times in purple. As such, attribute stimuli were presented on one-third of the trials and target stimuli on two-thirds of the trials. Each test block started with two additional attribute trials. In all blocks, stimuli were presented in a random order that was determined separately for each block and participant. Instructions about the upcoming task were presented before each block.

Each trial started with a fixation cross that was presented at the screen centre for 400 milliseconds. This was followed by the stimulus that stayed on the screen until a response was given. If participants made an incorrect response, a red X appeared on the screen for 400 milliseconds. The next stimulus appeared 1,200 milliseconds after a correct response or after the red X disappeared from the screen.

Results

Self-report questionnaire

Mean self-reported liking scores did not differ significantly from zero, either for VLD ($M = 0.02$, $SD = 1.35$) or for CD&V ($M = -0.15$, $SD = 1.35$), ts < 1. Both means did not differ, $t < 1$. In order to be able to compare self-reported liking with the IAT scores (which are inherently relative), we computed difference scores for the questionnaire items by subtracting self-reported liking of VLD from self-reported liking of CD&V. A positive score thus indicates a more positive attitude towards CD&V than towards VLD. Only 10 participants had previously voted in the past. Because of this small

number, we decided to omit this variable from the analyses. Six participants (12.50% of the total sample) indicated that they would (probably) vote for VLD during the next election and five (10.42%) indicated that they would (probably) vote for CD&V. Because of this small number, we again refrained from analysing these data further.

IAT

We calculated for each participant a D600 IAT score using the new scoring algorithm that Greenwald, Nosek, and Banaji (2003) recently recommended. The most important feature of this new scoring algorithm is that the IAT score for each participant is standardized and that the reaction time on trials with an incorrect response on each mixed test block is replaced by the mean latency for correct trials on that block, increased with a penalty of 600 milliseconds. We implemented the algorithm in such a way that a positive D600 score signifies a preference for CD&V over VLD. For ease of interpretation, we will, however, report only the mean untransformed reaction times, both for the IAT and the EAST tasks.

Inspection of the data showed that three participants could be regarded as outliers. One participant had an unusually high percentage of errors (where unusual is defined here and in the remainder of the paper as more than three standard deviation, above the mean of the total group). Two participants had an unusually high number of reaction times shorter than 300 milliseconds or longer than 3,000 milliseconds. The data of these participants were excluded from the analyses.[1]

As in the remainder of the paper, we assessed the reliability of our IAT scores using an odd–even split-half procedure. We first listed all trials in order of appearance, separately for each type of stimulus (i.e. VLD, CD&V, positive, negative), test block (VLD-positive, CD&V-positive) and participant. We then calculated for each participant a D600 score based on the odd trials of the two test blocks and a D600 score based on the even trials of the test blocks, thus making sure that each D600 score was based on an equal number of trials with each stimulus type. We then calculated the split-half reliability by correlating these two scores and applying a Spearman-Brown correction. Using this procedure, we found in this experiment a split-half reliability of .88 for the D600 measure.

Most importantly, the IAT D600 score correlated significantly with the difference score of the self-reported attitudes, $r = .37$, $p < .05$.[2] Overall, reaction times in the CD&V-positive task ($M = 741, SD = 153$) were somewhat shorter than reaction times in the VLD-positive task ($M = 760, SD = 154$). The D600 IAT score differed marginally significant from zero, $t(44) = 1.74, p = .09$, suggesting a slight overall preference for CD&V over VLD.

EAST

Analogous to Schmukle and Egloff (2006), we computed for each participant a relative D600 score in such a way that a more positive difference EAST score signifies a more

---

[1] *Greenwald et al. (2003) suggested to exclude only those participants from the analysis from whom more than 10% of their response latencies were shorter than 300 milliseconds. Following these suggestions led to the same conclusions for both the IAT and the EAST as the analyses presented in this article. Also, using D2SD IAT and EAST measures and IAT and EAST scores based on log-transformed latencies or errors instead of the reported D600 measures yielded similar conclusions.*

[2] *This correlation was obtained after partialling out the effects of task order (EAST first vs. IAT first), questionnaire order (questionnaire filled in before or after the reaction time tasks) and response assignment variations. All manipulated variables (including counterbalancing variables) were also statistically controlled for in all other correlations reported in this article that involved IAT or EAST scores. All significant correlations remained significant when the effect of the manipulated variables was not partialled out.*

positive attitude towards CD&V than towards VLD. We also calculated D600 scores for CD&V and VLD separately. For these separate D600 EAST scores, a more positive score signifies a more positive attitude towards the object. The data of one participant were removed from the analyses because of an unusually high number of trials on which the reaction time was shorter than 300 milliseconds or longer than 3,000 milliseconds. All relevant means can be found in Table 1.

Table 1. Mean untransformed reaction times and percentage of errors during the EAST as a function of target and response in Experiment 1

| | Response | | | |
| | Positive | | Negative | |
Target	M	SD	M	SD
CD&V				
Reaction time	628	127	642	141
Percentage of errors	4.19	4.67	5.12	5.23
VLD				
Reaction time	614	121	651	116
Percentage of errors	5.12	4.81	4.26	3.88

The odd–even split-half reliability of the different EAST scores was calculated in a similar way as the odd–even split-half reliability of the IAT score that we discussed above. The resulting split-half reliabilities were .35, .16 and $-.20$ for the CD&V D600 score, the VLD D600 score and the relative D600 score, respectively. The VLD D600 EAST score did not correlate with self-reported liking of VLD, $r = -.00$, or with liking of CD&V, $r = .09$. The CD&V D600 EAST score did not correlate with self-reported liking of CD&V, $r = .24$. There was a tendency for a correlation between the CD&V D600 score and self-reported liking of VLD, $r = .28$, $p = .07$. We also failed to find a correlation between the difference score of the self-reported attitudes and the relative D600 EAST score, $r = .09$.

A priori t tests showed that none of the D600 EAST scores differed significantly from zero, ps > .10, indicating that overall the participants had neutral attitudes towards the political parties. The D600 EAST scores for VLD and CD&V also did not differ from each other, $p > .10$.

Correlations between IAT and EAST scores
Since IAT scores are inherently relative, we computed the correlation between the IAT D600 score and the relative EAST D600 score. This correlation was not significant, $r = -.10$.

EXPERIMENT 2

In Experiment 2, we used the target items 'beer' and 'sprouts'. We chose these concepts because a pilot study showed that psychology students at Ghent University differed substantially with regard to their attitudes towards these concepts. We examined

whether IAT and EAST scores correlated with self-reported liking and consumption of beer and sprouts.

Method

Participants

The number of participants completing the IAT was 48 whereas 66 participants completed the EAST. The large majority of the participants were undergraduate psychology students at Ghent University. The remaining participants were students of other faculties at Ghent University. All participants were recruited by flyers that were distributed near classrooms. They received either €2 or €5 for their help and took part in groups of 4–12 people.

Measures and procedure

The experiment took place in the same dimly lit computer classroom where Experiment 1 was conducted. Before filling in an informed consent form, participants were told that the experiment consisted of a reaction time task on a personal computer and a short questionnaire. Half of the participants started by filling out a questionnaire. The other half received the questionnaire after the reaction time task was completed. As part of the questionnaire, participants rated on a seven-point scale (from − 3 to 3) how much they liked each target item. Also, participants indicated how often they typically ate sprouts or drank beer by using a six-point scale with the following response options: never, less than once a year, once a year, once a month, once a week, more than once a week. The IAT was identical to the IAT used in Experiment 1 except for the target stimuli, which were words and pictures referring to the categories sprouts and beer (see Appendix). The EAST was also identical to the EAST used in Experiment 1, except that we used the words SPRUITEN [sprouts] and BIER [beer] as target stimuli (see Appendix).

Results

Self-report questionnaire

The mean self-reported liking for sprouts did not differ significantly from zero ($M = 0.24$, $SD = 2.05$), $t(113) = 1.26$. Self-reported attitudes towards beer tended to be positive ($M = 0.31$, $SD = 1.94$), $t(110) = 1.69$, $p = .09$. Both means did not differ significantly from each other, $t < 1$. We computed difference scores for the questionnaire items by subtracting the responses on sprouts items from the responses on beer items. Thus, a positive score on the difference scores reflect a more positive attitude towards beer than towards sprouts. Participants reported drinking beer more frequently compared with eating sprouts, $t(111) = −7.34$, $p < .001$.[3] In our sample, 33% reported drinking beer more than once a week, resulting in a ceiling effect for the beer consumption scale. In order to compare self-reported consumption behaviour with the relative IAT scores, we computed a

---

[3] *Since the consumption self-report items for beer and sprouts are actually on the ordinal measurement level, we also used non-parametric tests to analyse the consumption data. The results of these analyses yielded the same conclusions as the reported analyses, except for the marginally significant correlation between the difference consumption measures and the relative D600 EAST score, $\rho = −.23$, p = .08, which was, however, in the opposite direction of what one would expect.*

difference score for the behaviour items by subtracting the response on the sprouts item from the response on the beer item.

IAT

We calculated a D600 IAT score in such a way that a more positive score indicated a larger preference for beer over sprouts. The data of two participants were removed from the analysis either because of an unusually high percentage of errors or an unusually high number of reaction times less than 300 milliseconds or longer than 3,000 milliseconds.

The IAT D600 measure had an odd–even split-half reliability of .95. We found a strong correlation between the D600 IAT score on the one hand, and the difference score for the self-reported attitudes, $r = .52, p < .001$, and the difference score for the frequency of consumption question, $r = .42, p < .01$, on the other hand. The IAT D600 score was not significantly different from zero, $t(45) = 1.39, p = .17$, indicating that reaction times in the beer-positive task ($M = 715, SD = 155$) did not differ significantly from those in the sprouts-positive task ($M = 752, SD = 143$).

EAST

EAST D600 scores were calculated in the same way as in Experiment 1. Beer and sprouts scores were calculated so that a more positive EAST D600 score indicated a more positive attitude towards the target. We also calculated a relative EAST D600 score in a way that a positive relative EAST D600 score signifies a preference for beer over sprouts. The data of one participant were excluded because of an unusually high percentage of errors and the data of a second participant were discounted because she had an unusually high mean reaction time. The relevant means can be found in Table 2.

Table 2. Mean untransformed reaction times and percentage of errors during the EAST as a function of target and response in Experiment 2

| | Response | | | |
| | Positive | | Negative | |
Target	M	SD	M	SD
Beer				
Reaction time	635	106	632	93
Percentage of errors	4.30	4.77	4.25	4.25
Sprouts				
Reaction time	656	100	651	116
Percentage of errors	5.52	5.42	3.66	4.78

Odd–even split-half reliabilities of the various scores were .61, .61 and .63, for the relative D600 EAST score, the beer D600 EAST score and the sprouts D600 EAST score, respectively. The sprouts EAST D600 score and beer EAST D600 score did not correlate significantly with self-reported liking or consumption behaviour, $-.07 < r < .16$. The relative D600 EAST scores also did not correlate with the difference score of

self-reported liking, $r = -.03$, and the difference measure for consumption behaviour, $r = -.15$. None of the EAST D600 scores was significantly different from zero, all $ps > .10$.

EXPERIMENT 3

The known-group approach is a popular method for assessing the validity of measures. When applied to the measurement of attitudes, the underlying idea is that different groups of people, who can be assumed on *a priori* grounds to differ with regard to a certain attitude, should also differ with regard to their score on a valid measure of that attitude. This approach has been successfully used in many studies using the IAT. For example, Banse, Seise, and Zerbes (2001; also see Jellison, McConnell, & Gabriel, 2004) used this approach to validate the IAT as a measure of attitudes towards homosexuality and found significant differences in performance on a homosexuality IAT between homosexual and heterosexual participants. We conducted a conceptual replication of the IAT study by Banse *et al.* but also included an EAST measure of attitudes towards homo- and heterosexuality. In this manner, we could test the validity of both the IAT and the EAST and compare the measures with regard to the impact of group status.

Method

Participants

Ninety-one individuals (mean age 23.7 years, $SD = 7.15$) took part. We recruited participants by distributing flyers in which we asked for the help of heterosexual and homosexual men and women for a study on new methods to assess attitudes towards homosexuality. We also contacted several organizations for homosexuals, asking them to distribute our call for participants. Some of the participants were paid €5, others received a small gift or participated without something in return. Nine participants had to be excluded from the sample because they did not identify themselves as unambiguously heterosexual or homosexual (see the section Sexual orientation measure). Of the remaining 82 participants, 19 were heterosexual men, 20 were heterosexual women, 29 were gay men and 14 were lesbian women.

Measures and procedure

Before filling in an informed consent form, participants were told that the experiment consisted of two reaction time tasks on a personal computer and a questionnaire. For about half of the participants, the IAT was presented first. The other half received the EAST first. The questionnaire was always presented after the EAST and the IAT. The experiment was conducted individually or in groups of two and took place in the classroom described in Experiment 1 or in a dimly lit classroom containing one or two Pentium IV laptops with 15-inch screens. Eleven participants were tested in a room of the house of a local organization for homosexual youngsters.

Cognitive and affective attitudes towards homosexuality scale

As an explicit measure of attitudes towards homosexuality, we used a Dutch translation of the Skala Kognitive Einstellungen zur Homosexualität and the Skala Affektive Einstellungen zur Homosexualität (Seise, Banse, & Neyer, 2002). This questionnaire

consists of two scales: a cognitive attitude scale and an affective attitude scale. The cognitive scale comprises 18 statements about homosexuality (e.g. 'gay couples should have the right to get married'). Participants are asked to indicate on a five-point Likert scale to what extent they agree with each of these statements. The affective scale is composed of 18 items describing situations or events relating to homosexuality (e.g. 'My daughter tells me she's a lesbian'). Participants indicate on a five-point Likert scale to what extent they would feel comfortable in these situations. Negative items were reversed so that for all items, a higher score indicated a more positive attitude towards homosexuality. The scores on the cognitive and affective items were then summed separately to form a cognitive attitude score and an affective attitude score, respectively.

Sexual orientation measure

Analogous to the study of Banse *et al.* (2001), sexual orientation of participants was assessed by asking people to indicate their sexual identity (first item) and sexual and/or relational behaviour (second item) on a five-point scale ranging from 1 (exclusively heterosexual) to 5 (exclusively homosexual). The ratings on both items were averaged to form a sexual orientation scale. Following Banse *et al.*, participants with a mean score between 2 and 4 were regarded as not unambiguously homosexual or heterosexual and their data were excluded from the statistical analyses.

IAT

The IAT was identical to the one used in Experiment 1 except on the following points. The selection of the stimuli was based on the study of Banse *et al.* (2001). All attribute stimuli in the IAT were words. Target categories were 'homosexual' and 'heterosexual' and the target items consisted of two names, male or female, with a '+' sign between them, suggesting that these two names indicated a couple. For example, the stimulus 'THOMAS + ISABEL' (a male name + a female name) indicated a heterosexual couple whereas the stimulus 'ANNICK + ISABEL' (two female names) or 'THOMAS + PIETER' (two male names) indicated a homosexual couple. All attribute and target stimuli can be found in the Appendix.

The IAT consisted of (a) an attribute practice block of 24 trials, (b) a target practice block of 32 trials, (c) three test blocks of 32 trials each, (d) a reversed target practice block of 32 trials and (e) three reversed test blocks of, again, 32 trials each. During the attribute practice block, each attribute stimulus was presented four times. During the target practice blocks, each target stimulus appeared on two trials. During the test blocks, each attribute stimulus appeared on two trials and each target stimulus on one trial. All participants were instructed to press the left key for negative attribute words and the right key for positive attribute words. In the first series of three test blocks, target stimuli referring to homosexual couples were always assigned to the same key as negative attributes and heterosexual items were assigned to the same key as positive attributes (heterosexual-positive task). For all participants, the second series of test blocks corresponded to the homosexual-positive task (same key for homosexual and positive items and the other key for heterosexual and negative items).

EAST

Again, only the differences with Experiment 1 will be discussed. As attribute stimuli, we used the same stimuli as in the IAT (see Appendix). Targets were the items 'HOMO'

(Dutch for gay), 'LESBISCH' (Dutch for lesbian) and 'HETERO' (Dutch for hetero or straight). We selected these items because they represent the to-be-measured concepts in an unambiguous manner. We also presented the string '#####' as a baseline item so that we could correct for possible baseline differences in the ease of emitting the two responses. For instance, some participants might be overall faster in responding with the right hand (i.e. the positive response) than with the left hand (i.e. the negative response) whereas the reverse might be true for another participant. These differences could, in principle, be estimated on trials with a neutral item such as a string of # signs and then used to correct each individual EAST score. However, given that adding this correction to the EAST scores did not change the pattern of results, we ignore these trials in the remainder of this paper. All targets were printed equally often in blue and in green. All participants were instructed to press the left key for negative attributes and green targets and the right key for positive attributes and blue targets. The EAST consisted of (a) an attribute practice block of 32 trials (each attribute stimulus presented four times), (b) a target practice block of 24 trials (each target stimulus presented six times) and (c) five mixed test blocks of 48 trials each (each attribute stimulus presented twice and each target stimulus presented eight times). We also manipulated the response–stimulus interval. About half of the participants received a version of the EAST in which the interval was 400 milliseconds (similar to the IAT) whereas the interval was 1,200 milliseconds for the other participants (similar to the EAST in the previous experiments). We will not discuss this variable further because it did not have a significant effect on the results.

Results

Cognitive and affective attitudes towards homosexuality scale and sexual orientation measure

The internal consistency of the attitude scales and the sexual orientation measure was high, with a Cronbach's alpha of .86 for the cognitive scale, .94 for the affective scale and .97 for the sexual orientation measure. The Pearson correlation between the two attitude scales was significant, $r = .57$, $p < .001$. We subjected the attitude questionnaire data to an ANOVA with type of attitude (cognitive vs. affective) as a within-subjects variable and sexual orientation (heterosexual vs. homosexual) as a between-subjects variable. The main effect of type of attitude was significant, $F(1, 80) = 373.56$, $p < .001$, showing that cognitive attitudes were more positive than affective attitudes. We also found a main effect of sexual orientation group, $F(1, 80) = 32.87$, $p < .001$, showing that homosexual participants had more positive attitudes towards homosexuality than heterosexual participants. Both variables interacted, $F(1, 80) = 7.95$, $p < .01$. Independent sample t tests showed that homosexual participants had more positive cognitive attitudes than heterosexual participants, $t(80) = 4.00$, $p < .001$, $M = 4.67$, $SD = 0.23$ and $M = 4.28$, $SD = 0.59$, for homo- and heterosexual participants, respectively. The difference in affective attitudes was even larger, $t(80) = 5.61$, $p < .001$, $M = 3.67$, $SD = 0.49$ and $M = 2.94$, $SD = 0.68$, for homo- and heterosexual participants, respectively.

IAT

We computed a D600 IAT score in the same way as in previous experiments. A positive IAT D600 score reflects a positive attitude towards homosexuality. The data of four

participants were removed from the analysis (one on the basis of an unusually high error percentage, one based on unusually long reaction times, one based on an unusual number of reaction times below 300 milliseconds or above 3,000 milliseconds and one based on both an unusually high percentage of errors and number of reaction times below 300 milliseconds or above 3000 milliseconds). One other participant failed to complete the IAT. Of the remaining participants, 39 were homosexual and 38 were heterosexual.

The odd–even split-half reliability of the D600 IAT score was .83. The IAT D600 score correlated significantly with the cognitive attitude scale, $r = .29$, $p < .05$, and with the affective attitude scale, $r = .57$, $p < .001$. When we used Cohen and Cohen's (1983) formula to compare the correlations for the cognitive and affective scales, we found that, compared with the cognitive scale, the affective scale was more strongly correlated with the IAT D600 score, $t(71) = 3.39$, $p < .01$. Independent sample t tests showed that homosexual participants had a larger IAT D600 score than heterosexual participants, $t(75) = 5.12$, $p < .001$, thus revealing more positive attitudes towards homosexuality. Homosexual participants were faster in the homosexual-positive task ($M = 1,030$, $SD = 232$) than in the heterosexual-positive task ($M = 1,099$, $SD = 230$) and thus had positive D600 IAT scores, $t(38) = 2.86$, $p < .01$. Heterosexual individuals were faster in the heterosexual-positive task ($M = 1,004$, $SD = 193$) than in the homosexual-positive task ($M = 1,113$, $SD = 238$) and thus had negative IAT D600 scores, $t(37) = -4.43$, $p < .001$.

EAST

The D600 EAST scores were calculated in the same way as in previous experiments. In order to increase comparability with the IAT, the reaction times and error percentages for the trials on which 'gay' or 'lesbian' appeared were combined into a 'homosexual' category.[4] Thus, we computed EAST scores for 'homosexual' and 'heterosexual' in such a way that a more positive EAST score indicated a more positive attitude towards the target category. We also calculated a relative D600 EAST score in a way that a positive score signifies a preference for homosexual over heterosexual. The data of five participants were excluded from the analyses (two on the basis of error percentage, one because of mean reaction time, one on the basis of number of trials below 300 milliseconds or above 3,000 milliseconds, one on the basis of both number of trials below 300 milliseconds or above 3,000 milliseconds and error percentage). Another participant reported to be colour blind and did not perform the EAST. Of the remaining participants, 40 were homosexual and 36 were heterosexual. Table 3 contains the relevant mean reaction times and error data.

The odd–even split-half reliability was .02, .43 and .25 for the relative D600 EAST score, the D600 EAST score for homosexuality and the D600 EAST score for heterosexuality, respectively. None of the EAST D600 scores correlated significantly with the cognitive scale of the questionnaire, $-.16 < r < .18$, or with the affective scale of the questionnaire, $-.15 < r < .18$

We analysed the D600 EAST scores using an ANOVA with group (homosexual vs. heterosexual) as a between-subjects variable and target category (homosexual vs.

---

[4] Note, however, that when we did not average over these two categories, analyses revealed the same (null) effects and (near zero) correlations for each category (i.e. gay and lesbian).

Table 3. Mean untransformed reaction times and percentage of errors during the EAST as a function of group, target and response in Experiment 3

	Response			
	Positive		Negative	
Target	*M*	*SD*	*M*	*SD*
Heterosexuals				
Hetero				
Reaction time	751	165	805	185
Percentage of errors	3.75	5.90	3.19	3.99
Homo				
Reaction time	762	173	773	136
Percentage of errors	3.47	4.56	2.99	3.77
Homosexuals				
Hetero				
Reaction time	788	172	814	174
Percentage of errors	1.28	2.49	1.79	3.14
Homo				
Reaction time	780	166	791	157
Percentage of errors	1.86	2.79	2.95	3.34

heterosexual) as a within-subjects variable. This analysis did not reveal an interaction between target category and group, or any other effect, all Fs < 1.17. Thus, we failed to find group differences in EAST performance.

Correlations between the EAST and the IAT

As was the case in Experiment 1 (but not Experiment 2), each participant completed both an IAT and an EAST. However, the relative D600 EAST score did not correlate with the D600 IAT score, $r = .02$.

GENERAL DISCUSSION

The results are clear: the IAT consistently outperformed the EAST as a measure of inter-individual differences in attitudes. Whereas IAT scores correlated in the expected manner with self-reported attitudes (Experiments 1–3) and self-reported behaviour (Experiments 2 and 3), we obtained no evidence that EAST scores were related to these criteria. Also, whereas the split-half reliability of IAT scores was consistently high across the three experiments (range of .83 to .95), split-half reliability of the EAST scores was consistently lower (range of $-.20$ to .63). Together, these results lead to the conclusion that the IAT is a better measure of inter-individual differences in attitudes than the EAST, at least for the adaptations that were explored in the present studies.

However, our studies have a number of limitations that force us to qualify this conclusion. First, we measured attitudes towards only a limited set of attitude objects. It is possible that the EAST fails to measure inter-individual differences in these attitudes but not differences in attitudes towards other attitude objects. This argument is, of course, difficult to refute. We can only point out that we see no obvious reason why

the EAST fails to measure the attitudes on which we focused and why it would be more successful in measuring other attitudes. For instance, De Houwer (2003b) pointed out that the EAST cannot be expected to measure inter-individual differences when there are no clear inter-individual differences. Because of this, we purposefully selected attitude objects for which there were clear inter-individual differences in attitudes. The fact that we were successful in selecting such attitude objects was also evidenced by strong correlations between IAT scores and criterion variables. It is thus unclear to us why the EAST would be particularly insensitive to the attitudes on which we focused.

Second, we used only a limited set of criterion variables. That is, the validity of the EAST was indexed by examining whether EAST scores were related to self-reported attitudes and behaviour. One could argue that it is odd to use correlations with such explicit measures as a criterion for assessing the validity of implicit measures (but see Greenwald *et al.*, 2003, for a similar approach). Implicit and explicit measures are assumed to capture different types of attitudes (e.g. Wilson, Lindsey, & Schooler, 2000) or the same attitude under different conditions (e.g. De Houwer, in press; Fazio & Olson, 2003). Hence, implicit and explicit measures should diverge, at least to a certain extent. However, Nosek (2005) recently provided theoretical arguments and empirical findings to back up the claim that implicit and explicit measures can be strongly related under certain conditions, for instance when self-presentation concerns are low, attitudes are strong, and attitudes are thought to be distinctively personal rather than normative. These conditions seem to hold for the attitude objects that we selected in Experiments 1 and 2 where implicit–explicit correlations were the main validity criterion. For attitudes towards mainstream political parties (Experiment 1) and food items (alcohol and sprouts; Experiment 2), there is little reason to think that self-presentation concerns would be strong. Also, political and food preferences are highly personal and often strong. Moreover, the fact that IAT scores were related to the criterion variables suggests that the lack of correlation with the EAST scores was not a result of the nature of the criterion variables. Hence, we believe that our criterion variables did allow us to assess in a fair manner the validity of the EAST as a measure of inter-individual differences in attitudes. Nevertheless, we cannot exclude the possibility that the EAST might prove valid when related to other criteria.

Our studies were also limited in that we used only a few of the possible procedural implementations of the EAST. First, in all our experiments, the EAST involved only a limited number of trials. For instance, each reaction time EAST score in Experiments 1 and 2 corresponded to the mean reaction time on 32 trials (e.g. with the target beer and a positive response) minus the mean reaction time on 32 other trials (e.g. with the target beer and a negative response). Given that reaction times are notoriously sensitive to a variety of sources of error variance (e.g. trial order effects, variations in concentration), it is likely that a reliable estimate of an attitude can be obtained only when the means and thus the EAST score are based on more observations. This explanation is in line with the findings of Voss, Rothermund, and Wentura (2003). Using a task very similar to the EAST, they examined differences in attitudes towards a variety of objects. The EAST(-like) score for each object was determined by averaging across the data of many participants. Hence, each score was based on a large number of trials. In this case, scores were valid (i.e. corresponded with normative data and experimentally induced attitudes) and reliable. It is thus possible that one could increase the reliability and validity of the EAST as a measure of inter-individual differences in attitudes by increasing the number of relevant trials. Note, however, that this would make the measure more

time consuming, especially when the aim is to measure several attitudes within a single EAST.

A second procedural limitation of the present studies is that, in line with the studies of De Houwer (2003b), we presented only verbal stimuli during the EAST and asked participants to respond on the basis of the colour of the target stimuli. As De Houwer (in press) pointed out, an EAST score can reflect the attitude towards a target concept only if and to the extent that (a) the presented target stimulus is processed, (b) the to-be-measured attitude is activated by this stimulus and (c) activation of this attitude influences performance. It might be the case that these conditions are not fulfilled in an optimal manner when the target stimuli are words and colour is the relevant response feature. For instance, there is evidence that, compared with words, pictorial stimuli more directly activate the associated attitude in memory (e.g. De Houwer & Hermans, 1994). Hence, EAST scores might be reliable and valid measures of attitudes if pictorial rather than verbal stimuli are used. In support of this argument, Huijding and de Jong (2005) found that an EAST with picture stimuli but not an EAST with verbal stimuli was related to inter-individual differences in fear of spiders and avoidance of actual spiders. Likewise, using a pictorial EAST, we found that an EAST measure of attitudes towards meat and vegetables differentiated between vegetarians and meat eaters, be it to a lesser extent than an IAT measure of those attitudes (De Houwer & De Bruycker, in press).

If one does present words as target stimuli, there are several possible ways to increase the likelihood that the target words are processed and do activate the to-be-measured attitude in memory. First, one can ask participants to respond on the basis of a semantic feature of the target words (e.g. the semantic category to which they belong) rather than on the basis of their colour. Huijding and de Jong (in press) followed this approach and found that the resulting EAST measure of attitudes towards spiders uniquely predicted spontaneous behaviour towards spiders over and above what was predicted on the basis of a self-report measure and an IAT attitude measure. Second, one can pre-activate the target concept so that target words are more likely to activate this concept. For instance, Ellwart, Becker, and Rinck (2005) presented target words that could be interpreted as referring to spiders (e.g. web, net). Before the start of the EAST, Ellwart and colleagues manipulated the activation level of the concept 'spider' by asking participants to look at spider-related or unrelated pictures. EAST scores were different for high-spider fearful than for low-spider fearful participants, but only after the concept 'spider' had been activated. Third, one can encourage participants to process the target words semantically by making their meaning task relevant. In a study conducted at our laboratory (De Houwer & De Bruycker, 2007), we presented positive adjectives, negative adjectives and the target words 'beer' and 'sprouts'. All words were presented as often in upper case as in lower case letters. Participants were asked to press one key for positive words (i.e. the positive response) and another key for negative words (i.e. the negative response). However, the task was different when the word 'beer' or 'sprouts' was presented. When they saw the word 'beer' or 'sprouts', they had to pay attention to the letter case and press the positive key if the word was presented in upper case letters and the negative key if it appeared in lower case letters. Thus, participants had to identify the presented word before they could know whether valence or letter case was response relevant. We therefore referred to this variant as the identification or ID-EAST. Results showed that students who typically consume more than 10 units of alcohol per week had a more positive attitude towards beer as measured by the ID-EAST than students who typically consume less than 3 units of alcohol per week. When taking the results of these four studies together (De Houwer & De Bruycker, 2007; Ellwart *et al.*,

2005; Huijding & de Jong, 2005, in press), it seems that some implementations of the EAST can be used to successfully measure inter-individual differences in attitudes. The key element seems to be that the task needs to be set up in such a way that participants are encouraged to process the target stimuli and that the to-be-measured attitude is activated.

Whereas the results that were reported in this paper raise doubts about the reliability and validity of (certain implementations of) the EAST as a measure of inter-individual differences in (certain) attitudes, they do provide further evidence for reliability and validity of the IAT as a measure of inter-individual differences in attitudes. Our results replicate and add to previous evidence showing that differences in IAT scores are related in a meaningful manner to differences in self-reported attitudes and behaviour (see Nosek *et al.*, 2007, for a review). One aspect of our IAT results is particularly noteworthy. In Experiment 3, IAT scores were related more to the affective component than to the cognitive component of attitudes towards homosexuality. This observation is in line with the suggestion that the IAT captures spontaneous evaluative associations with the attitude object (e.g. Greenwald & Banaji, 1995; Wilson *et al.*, 2000).

In summary, our results suggest that the EAST, as introduced by De Houwer (2003b), does not provide an adequate measure of inter-individual differences in attitudes. This does not exclude the possibility that the EAST might provide a valid and reliable measure of attitudes at the supra-individual level (see De Houwer, 2003b; Voss *et al.*, 2003, for supporting evidence) or that certain variants of the EAST can provide valid measures of inter-individual differences in attitudes. Our results should also not discourage researchers from continuing to look for alternatives for the IAT. These results do not take away the inherent limitations of IAT, such as the relative nature of IAT scores, nor do they disprove claims that IAT effects might be biased by non-attitudinal factors (e.g. De Houwer, 2002; Rothermund & Wentura, 2004). Moreover, regardless of the possible benefits and shortcomings of the IAT, it would always be good to have an alternative task that could be used independently to verify findings that have been obtained with the IAT (e.g. Nosek & Banaji, 2001). What the present results strongly suggest, however, is that the EAST, as described by De Houwer (2003b), does not offer the alternative measure that we were looking for.

Acknowledgements

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.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression correlation analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Houwer, J. (2001). A structural and process analysis of the implicit association test. *Journal of Experimental Social Psychology*, *37*, 443–451.

De Houwer, J. (2002). The implicit association test as a tool for studying dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behaviour Therapy and Experimental Psychiatry*, *33*, 115–133.

De Houwer, J. (2003a). A structural analysis of indirect measures of attitudes. In J. Musch & K.C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219-244). Mahwah, NJ: Erlbaum.

De Houwer, J. (2003b). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77-85.

De Houwer, J. (in press). Comparing measures of attitudes at the procedural and functional level. In R. Petty, R.H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures*. Mahwah, NJ: Erlbaum.

De Houwer, J., & De Bruycker, E. (in press). Differences in implicit attitudes toward meat and vegetables in vegetarians and omnivores. *International Journal of Psychology*.

De Houwer, J., & De Bruycker, E. (2007). The identification-EAST as a valid and reliable measure of attitudes toward alcohol-related stimuli. *Journal of Behaviour Therapy and Experimental Psychiatry, 38*, 133-143.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12*, 45-61.

De Houwer, J., & Hermans, D. (1994). Differences in the affective processing of words and pictures. *Cognition and Emotion, 8*, 1-20.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp.179-194). New York: Guilford Press.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about questions: Meaning, memory, expression and social interaction in surveys* (pp. 204-237). New York: Sage.

Ellwart, T., Becker, E. S., & Rinck, M. (2005). Activation and measurement of threat associations in fear of spiders: An application of the extrinsic affective Simon task. *Journal of Behaviour Therapy and Experimental Psychiatry, 36*, 281-299.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

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

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Huijding, J., & de Jong, P. J. (2005). A pictorial version of the extrinsic affective Simon task. *Experimental Psychology, 52*, 289-295.

Huijding, J., & de Jong, P. J. (in press). Specific predictive power of implicit associations for automatic fear behaviour. *Behaviour Research and Therapy*.

Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: Ingroup preferences and related behaviours and beliefs among gay and straight men. *Personality and Social Psychology Bulletin, 30*, 629-642.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565-584.

Nosek, B., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic Processes in Social Thinking and Behaviour* (pp. 265-292). Philadelphia, PA: Psychology Press.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the implicit association test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.

Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2005). *Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity.* Manuscript submitted for publication.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139–165.

Schmukle, S. C., & Egloff, B. (2006). Assessing anxiety with extrinsic Simon tasks. *Experimental Psychology*, *53*, 149–160.

Seise, J., Banse, R., & Neyer, F. J. (2002). Individuelle Unterschiede in impliziten und expliziten Einstellungen zur Homosexualität. *Zeitschrift für Sexualforschung*, *15*, 21–42.

Teige, S., Schnabel, K., Banse, R., & Asendorpf, J. B. (2004). Assessment of multiple implicit self-concept dimensions using the extrinsic affective Simon task (EAST). *European Journal of Personality*, *18*, 495–520.

Voss, A., Rothermund, K., & Wentura, D. (2003). Estimating the valence of single stimuli: A new variant of the affective Simon task. *Experimental Psychology*, *50*, 865–896.

Wigboldus, D., Holland, R. W., & van Knippenberg, A. (2005). Single target implicit associations. Manuscript submitted for publication.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126.

Appendix

List of items used in the experiments

Note. Translation of the Dutch words is presented within brackets. Words printed in upper case letter were presented in upper case letters, those printed in lower case letters were presented in lower case letters.

Experiment 1: CD&V VLD IAT

- Positive attributes: picture of a butterfly, picture of a little seal, GOED [good] and eerlijk [honest]
- Negative attributes: picture of a cockroach, picture of a human skull, SLECHT [bad] and gemeen [mean]
- CD&V: political party's logo, an orange square (orange is the colour associated with CD&V), CD&V and cd&v
- VLD: the polical party's logo, a blue square (blue is the colour associated with VLD), VLD and vld

Experiment 1: CD&V VLD EAST

- Positive attributes: VLINDER [butterfly], ZEEHONDJE [little seal], GOED [good], EERLIJK [honest]
- Negative attributes: KAKKERLAK [cockroach], SCHEDEL [skull], SLECHT [bad], GEMEEN [mean]
- CD&V: CD&V
- VLD: VLD

Experiment 2: Sprouts beer IAT

- Positive attributes: picture of a butterfly, picture of a little seal, GOED [good] and eerlijk [honest]

- Negative attributes: picture of a cockroach, picture of a human skull, SLECHT [bad] and gemeen [mean]
- Sprouts: two pictures of sprouts, SPRUITEN [sprouts], spruiten [sprouts]
- Beer: two pictures of beer, BIER [beer], bier [beer]

Experiment 2: Sprouts beer EAST

- Positive attributes: VLINDER [butterfly], ZEEHONDJE [little seal], GOED [good], EERLIJK [honest]
- Negative attributes: KAKKERLAK [cockroach], SCHEDEL [skull], SLECHT [bad], GEMEEN [mean]
- Sprouts: SPRUITEN [sprouts]
- Beer: BIER [beer]

Experiment 3: Homosexuality IAT

- Positive attributes: PLEZIER [fun], COMPLIMENT [compliment], GOED [good], EERLIJK [honest]
- Negative attributes: ELLENDE [misery], BELEDIGING [insult], SLECHT [bad], GEMEEN [mean]
- Homosexual: ANN + MIA, NELE + ILSE, SOFIE + HEIDI, ANNICK + ISABEL, WIM + TOM, ERIK + KOEN, JONAS + DAVID, THOMAS + PIETER
- Heterosexual: WIM + MIA, ANN + TOM, ERIK + ILSE, NELE + KOEN, JONAS + HEIDI, SOFIE + DAVID, THOMAS + ISABEL, ANNICK + PIETER

Experiment 3: Homosexuality EAST

- Positive attributes: PLEZIER [fun], COMPLIMENT [compliment], GOED [good], EERLIJK [honest]
- Negative attributes: ELLENDE [misery], BELEDIGING [insult], SLECHT [bad], GEMEEN [mean]
- Targets: HOMO [gay], LESBISCH [lesbian], HETERO [hetero], #####