

A MODIFIED EXTRINSIC AFFECTIVE SIMON TASK (EAST) TO ASSESS THE AFFECTIVE VALUE OF PICTORIAL STIMULI: NO INFLUENCE OF AGE AND EDUCATIONAL LEVEL

Jorg HUIJDING & Peter DE JONG

Erasmus University & University of Groningen, The Netherlands

There is a range of concepts that are relevant to emotion research that may not be readily captured in words. Therefore, this experiment tested whether the Extrinsic Affective Simon Task (EAST: De Houwer, 2003) can be modified to assess automatic affect using pictorial stimuli. In addition, since the EAST is a relatively complex task, we tested whether the EAST can also be successfully employed in community samples. Participants ($n = 60$) who varied considerably in age and educational level completed a pictorial version of the EAST. Results show that the pictorial EAST is sensitive to the valence of pictorial stimuli and has an acceptable reliability. However, this pattern of results emerged most clearly in the error data. A task-shifting account for these findings is discussed. The findings underline the potential usefulness of the EAST as a tool for assessing automatic affect, but urge caution with interpreting data that include high error rates.

Introduction

During the last decade several measures have been developed to assess automatic affective associations. The major appeal of such measures is that they may be relatively robust against social desirability concerns and offer the opportunity to tap information that is not (necessarily) available for conscious introspection (e.g., Fazio & Olson, 2003). One of the most widely used measures of automatic associations is the Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998). The task has good psychometric properties, is flexible, and produces relatively large effect sizes (Greenwald & Nosek, 2001). However, inherent to its design the IAT can only be used to assess relative associations for bipolar target concepts (e.g., ingroup vs. out-

Jorg Huijding is Dr. at the Department of Clinical and Health Psychology, Erasmus University, Rotterdam, The Netherlands; Peter de Jong is Professor at the Department of Clinical and Developmental Psychology at the University of Groningen, The Netherlands.

The experiment reported in this paper was conducted at the Department of Medical, Clinical, and Experimental Psychology of Maastricht University, Maastricht, The Netherlands.

The author wishes to thank Bert Hoekzema for his technical assistance, and Monique Nijssen for her help in designing the experiment and collecting the data.

Correspondence concerning this article should be addressed to Jorg Huijding, Erasmus University, Institute of Psychology, Department of Clinical and Health Psychology, P.O. BOX 1738, 3000 DR Rotterdam, The Netherlands. E-mail: Huijding@fsw.eur.nl

group; men vs. women). This renders this task suboptimal for unipolar concepts that have no immanent, meaningful contrast such as “self”, “smoking” (e.g., Huijding, de Jong, Wiers, & Verkooijen, 2005), or feared stimuli such as spiders (de Jong, van den Hout, Rietbroek, & Huijding, 2003), etc.

Interestingly, De Houwer recently developed a measure for indirectly assessing specific (automatic) affective associations with a *single* target category without requiring a contrast category: the Extrinsic Affective Simon Task (EAST: De Houwer, 2003; cf. De Houwer & Eelen, 1998). De Houwer (2003) instructed participants to sort stimuli that appeared in the middle of a computer screen as fast as possible using a left and a right response key. There were two types of stimuli: attribute words, that were presented in white colour, and target words, that were presented in green or blue. The attribute stimuli consisted of positive and negative words, that had to be sorted on the basis of their valence. The target words were positive and negative words that were presented equally often in green and blue, and had to be sorted on the basis of their colour. Throughout the task each response key was the correct response for either the positive or negative attribute stimuli resulting in a positive and a negative response key. During the critical phase of the task, participants had to sort the attribute and the target words simultaneously. In effect, the correct response for each target stimulus was for half of the trials pressing the positive, and for the other half pressing the negative response key. De Houwer (2003) found that during the critical phase, participants performed better when the intrinsic stimulus valence and the extrinsic response valence were congruent than when they were incongruent. Thus, although the valence of the target stimuli should be ignored, it did influence task performance. Task performance may thus yield information concerning the valence of target stimuli. Initial findings with the EAST suggest that the EAST may be an efficient tool for assessing affective evaluations of a wide range of (unipolar) concepts, including self-esteem (e.g., De Houwer, 2003), spiders (Ellwart, Becker, & Rinck, 2005; Huijding & de Jong, 2006), and alcohol (De Houwer, Crombez, Koster, & De Beul, 2004; de Jong, Wiers, & van de Braak, & Huijding, in press).

Until now, studies using the EAST typically relied on verbal stimuli. Recently, however, Huijding and de Jong (2005) designed a pictorial variant of the EAST, and found preliminary evidence indicating that this pictorial EAST is sensitive to the affective value of normatively valenced stimuli. This adds to the flexibility of the EAST and opens up the opportunity to assess concepts that may not be readily captured in words. These include for instance concepts that are relevant in the context of phobic fear and anxiety (e.g., spiders, disapproving faces), stereotyping (e.g., physical features), and addiction research (e.g., smoking, drinking). In a similar vein, pictorial stimuli might be useful for the assessment of the implicit affective value of

(un)conditioned stimuli in the context of evaluative and aversive conditioning (e.g., Vansteenwegen, Crombez, Bayens, & Eelen, 1998). In addition, pictorial stimuli may sometimes provide ecologically more valid exemplars (e.g., spiders, angry face). Furthermore, because pictorial stimuli are argued to have more direct access to semantic information than verbal stimuli do (e.g., De Houwer & Hermans, 1994; De Houwer & d'Ydewalle, 1994; Glaser & Glaser, 1989), a pictorial EAST might be more sensitive to individual differences than a verbal EAST. An important additional advantage of a pictorial EAST is that it can be used to assess participants that cannot read or cannot read fluently, like (young) children, individuals with a low IQ, or individuals that speak a different language.

The present study aimed to test whether the pictorial EAST-effects are robust, and, perhaps even more importantly, whether similar results can be obtained in community samples. That is, the initial findings with the pictorial EAST were based on a sample of college students (Huijding & de Jong, 2005). Since the pictorial EAST is a relatively complex task (compared to the IAT), it is well conceivable that it is less suited for non-student (e.g., clinical) samples, or other samples with a lower level of education or of older age. To explore this important issue, we tested whether the initial findings with a pictorial version of the EAST that were reported by Huijding and de Jong (2005) could be replicated in a non-student, community sample that varied considerably in age and educational level.

Method

Participants

Participants were recruited following a "snowball" procedure in which the experimenter asked individuals from her personal environment (family, friends, neighbours, colleagues), who in turn asked their family, friends and colleagues, and so forth. This resulted in an unselected, heterogeneous community sample that comprised of 60 women who participated in return for a small gift. Due to problems logging the data, one participant had to be excluded from all analysis. The final sample varied considerably in age ($M = 41.6$, $SD = 15.2$, range = 14-76) and educational level. Educational level was based on the different levels in the Dutch educational system, and was defined as the highest form of education the participants successfully completed, ranging from 0 = no education completed to 10 = masters degree ($M = 6.3$, $SD = 2.2$). Two participants failed to report their educational level, one of whom also failed to report her age.

EAST

The pictorial EAST is formally similar to the verbal EAST described in the introduction. However, instead of words, participants were presented with pictorial target and attribute stimuli. Similar to the verbal EAST, the attribute stimuli had to be sorted on the basis of their valence. Target stimuli had to be sorted on the basis of their form. The task consisted of 3 phases: (1) practice sorting attribute pictures on the basis of their valence; (2) practice sorting target pictures on the basis of their form; (3) critical combined sorting of target and attribute pictures. Form was chosen as the relevant feature because this integrates the task-irrelevant feature (e.g., picture content) and the task-relevant feature (e.g., picture form). If the relevant and irrelevant feature are not integrated the picture content might be too easily ignored, reducing interference effects (e.g., Kindt & Brosschot, 1999). For this reason recent experiments in the context of emotional Stroop research employed a colour filter over the pictorial stimuli as the relevant feature (e.g., Constantine, McNally, & Hornig, 2001), rather than using stimuli consisting of separate relevant (e.g., a coloured dot or frame) and irrelevant (e.g., a picture of a spider) task features (e.g., Lavy & van den Hout, 1993). However, using a colour filter may undermine the strength (and/or validity) of the task-irrelevant stimulus feature. For example, a spider fearful participant may not fear a yellow spider, and a green disapproving face may be funny, rather than distressing to a social anxious individual.

Following this, we decided to use form (i.e., portrait or landscape) as the relevant stimulus feature for the target stimuli. All target stimuli were oblong pictures that were presented in either 'portrait' or 'landscape' format (see Appendix). The attribute stimuli were all square pictures with a yellow border around it. This border was included to facilitate discrimination between the square and the oblong pictures. There were five categories of target pictures; positive, neutral, negative, spider, and white. The spider and white pictures were included for exploratory reasons. The attribute pictures were clearly positive or negative. All positive, negative, and neutral pictures were selected from the International Affective Picture System (IAPS: Lang, Bradley, & Cuthbert, 1996) on the basis of valence and arousal (high arousal for positive and negative pictures, low arousal for neutral pictures).¹ Each

¹ The specific IAPS pictures that were used as target and attributes were as follows: positive square (1750, 2150, 2550, 5910, 8501); negative square (3063, 3080, 3130, 3500, 6313); positive oblong (1710, 2050, 8190, 8490, 8496); negative oblong (3000, 3053, 3170, 6350, 9410), and; neutral oblong (2190, 7004, 7006, 7010, 7175). Note that negative pictures are generally rated as more arousing than positive pictures, with the exception of positive pictures with a sexual content. Since we deemed sexual pictures inappropriate for the present experiment, selecting positive and negative pictures of a similar valence led the negative pictures to be more arousing than the positive pictures. The spider pictures were selected from previous research on spider fear at Maastricht University (e.g., Huijding & de Jong, 2006). The white pictures were white squares that were programmed into the EAST in MEL professional (Schneider, 1989).

attribute and target category consisted of 5 exemplar pictures. During Phase 1, each square yellow-bordered picture was presented 3 times (30 trials). In Phase 2, each oblong picture was presented once in 'portrait' and once in 'landscape' format (50 trials). During the third phase each square picture was presented 12 times and each oblong picture was presented 4 times in 'portrait' and 4 times in 'landscape' format, with each size appearing equally often in the portrait and landscape exemplars of each category (320 trials).

To prevent participants from focusing on one point of the screen while discriminating between portrait and landscape pictures (limiting the processing of the picture content), oblong pictures were presented in 5 different sizes (also see Appendix). The response requirements (left key, right key) for the targets (portrait, landscape) and the attributes (positive, negative) were counterbalanced over participants using four task versions. Each phase was preceded by specific response instructions. Following a correct response, the stimulus was immediately replaced by a fixation dot in the middle of the screen that remained there for 500 ms. Then the next stimulus was presented. If after 2500 ms no response was given, or if a given response was incorrect, the Dutch word "fout" (error) appeared briefly (250 ms) above the stimulus. In the mean time the stimulus remained on the screen until the participant gave the correct answer. The experiment was controlled by a MEL professional version 2 (Schneider, 1989) that was implemented on an IBM compatible 133 MHz Pentium personal computer, with a 14-inch monitor.

Procedure

All participants were tested individually. After completing the pictorial EAST participants filled in the questions concerning age and educational level. They then received the small gift (an incense candle).

Results and discussion

Prior to statistical analyses, correct responses faster than 300 ms (0.2%) were recoded to 300 ms (cf. De Houwer, 2003) (note that there were no responses slower than 3000 ms). Following this, the reaction times (RTs) of correct responses were log-transformed to normalise the skewed distribution (cf. De Houwer, 2003). Subsequently, separate EAST-scores for each target category were calculated for the RT as well as for the error (ER) data, by subtracting the mean RTs (or ERs) on target trials on which the correct response was positive from target trials on which the correct response was negative. Following this, a positive EAST-score indicates positive associations with the target category, whereas a negative EAST-score indicates negative associa-

tions. Overall, on 17.4% of the trials an error was recorded, which is a substantially higher error rate than usually reported for verbal EASTs (e.g., De Houwer, 2003; Ellwart et al., 2005) or for previous studies with the pictorial EAST (Huijding & de Jong, 2005). As can be seen in Table 1, the overall error rate was significantly correlated with age ($r = .29, p < .05$), suggesting that older participants made more errors than younger participants. The overall error rate was not significantly related to participants' level of education ($r = -.16, p > .2$). Eleven participants had an error rate over 30%. Although including the data of these participants did not lead to different results, they were removed from the analyses reported here. This was done because chances are high that participants with such high error rates did not understand the task instructions properly. Hence, their results may not reflect the associations we were interested in. Following this, the final sample consisted of 48 participants with a mean age of 39.6 years ($SD = 14.1$, range = 14 – 66), and a mean level of education of 6.5 ($SD = 2.3$, range = 1 – 10). Mean RTs and percentage of errors as a function of extrinsic response valence are shown in Table 1. The participants that were excluded from the analysis were significantly older than the included participants ($M = 49.8, SD = 17.5$, range = 23-76, $t(56) = -2.1, p < .05$), but had a similar level of education ($M = 5.7, SD = 1.8, t(55) = 1, p > .3$).

Previous research (e.g., Huijding & de Jong, 2005; De Houwer, 2003; De Houwer & Eelen, 1998) has shown that it is not unusual for Simon effects to occur in ERs. As can be seen in Table 1, the present results are in line with this. A 3 Target category (positive, neutral, negative) \times 4 Version ANOVA with repeated measures yielded the expected linear trend for Target in the ER data, $F(1, 44) = 48.6, MSE = 324.3, p < .01, \eta^2 = .53$. In addition a significant quadratic trend emerged, $F(1, 44) = 9.8, MSE = 120.9, p < .01, \eta^2 = .18$. Paired comparisons (with Bonferonni corrected alpha set to .016) showed that the EAST-score for positive was significantly more positive, $t(47) = 2.5, p < .05, d = 0.45$, and that of negative was significantly more negative than that for neutral, $t(47) = -6.8, p < .01, d = -1.26$. Additional simple t-tests showed that the EAST-score for positive was significantly more positive than 0, $t(47) = 2.8, p < .01$, that for negative was more negative than 0, $t(47) = -8.3, p < .01$, whereas that for neutral did not differ significantly from 0, $t(47) < 1$. For the white pictures and the pictures of spiders neither EAST-score differed significantly from 0, for both $t(47) < 1.2$. These results replicate the previous findings reported by Huijding and de Jong (2005) showing that the EAST can indeed be successfully adapted to assess the affective value of normative pictorial stimuli, and sustain the idea that the EAST can also be used in non-student community samples.

Table 1.
 RTs and ERs (sd between parentheses) as a Function of Response Valence, EAST-Scores, and Split-half Reliabilities
 for the EAST- Scores for each Stimulus Category.

Category	RT				ER			
	Extrinsic Response Valence		effect	α	Extrinsic Response Valence		effect	α
	positive	negative			positive	negative		
Positive	901 (209)	903 (166)	2 (159)	.72	10.5 (8.8)	16.7 (12.9)	6.2 (15.2)	.59
Neutral	894 (179)	908 (179)	14 (151)	.67	14.4 (11.3)	14.0 (11.2)	0.4 (13.7)	.60
Negative	997 (204)	938 (175)	-59 (165)	.50	28.2 (16.9)	8.7 (7.8)	-19.5 (16.3)	.54
Spider	862 (192)	876 (174)	14 (153)	.69	11.5 (12.1)	9.7 (10.4)	-1.8 (14.6)	.41
White	855 (174)	868 (182)	13 (118)	.61	9.1 (10.3)	11.7 (12.9)	2.6 (14.8)	.56

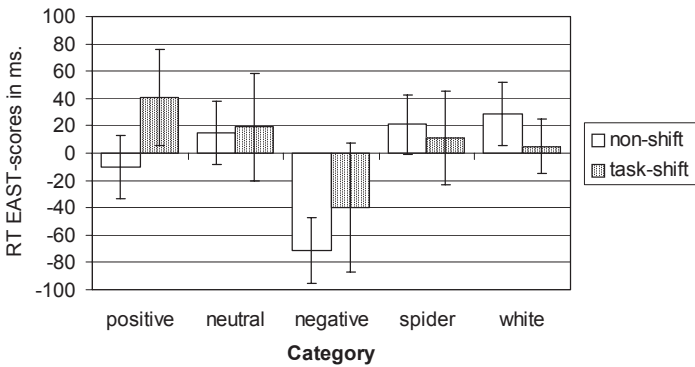
Note. Split-half reliabilities are based on EAST-scores. EAST-score for each category = RT or ER for negative response – RT or ER positive response. Cronbachs' alpha for the RTs is calculated on the basis of the log-transformed response latencies.

As can be seen in Table 1, the RT data showed much less consistent results. Similar to the ER data the RT EAST-scores were subjected to a 3 Target category (positive, neutral, negative) \times 4 Version ANOVA with repeated measures. Although this analysis yielded the expected linear trend, $F(1, 44) = 10.3$, $MSE = .01$, $p < .01$, $\eta^2 = .19$, also a significant quadratic trend emerged, $F(1, 44) = 7.1$, $MSE = 0.1$, $p < .05$, $\eta^2 = .14$. In addition, both the linear as well as the quadratic trends were qualified by a significant interaction with Version, $F(3, 44) = 4.6$, $MSE = .01$, $p < .01$, $\eta^2 = .24$, and $F(3, 44) = 3.1$, $MSE = .01$, $p < .05$, $\eta^2 = .176$, respectively. Paired comparisons showed that whereas the RT-based EAST-score for positive did not differ significantly from that for neutral, $t(47) < 1$, the EAST-score for negative was significantly lower than that for neutral, $t(47) = -3.4$, $p < .01$, $d = -.49$. Additional simple t-tests showed that the EAST-score for negative was significantly lower than 0, $t(47) = -2.5$, $p < .05$. The EAST-scores for positive and neutral did not differ significantly from 0, for both $t(47) < 1$. Also the EAST-scores for the spider and white pictures did not differ significantly from 0, for both $t(47) < 1$.

One explanation for the finding that in the present as well as in the previous pictorial EAST study (Huijding & de Jong, 2005) the expected pattern of results emerged much more clearly in the ER data than in the RT data may be found in the task-shifting account for EAST-effects (Voss, Schmitz, Teige, & Klauer, 2005). During the critical phase of the EAST, participants have to perform two tasks, sorting attribute trials on the basis of their valence, and sorting target trials on the basis of a non-evaluative relevant feature like word colour or picture form. The task-shift account poses that EAST-effects at least partly reflect participants' difficulty to shift between these two tasks. More specifically, if a participant has to sort a target trial just after having sorted an attribute trial, he or she will have to shift from using an evaluative sorting rule to a non-evaluative sorting rule (e.g., form or colour). It may well be that initially both task rules are activated simultaneously. It will then require less effort to select the correct response key if both rules point to the same key than when they point to a different response key. In effect, an EAST-effect will emerge which reflects faster responses on congruent than on incongruent trials. Participants may also fail to shift between tasks. If they do so, they will only make a mistake when the valence of the correct response key is incongruent with their affective association with the target stimulus. It is likely that because in the pictorial EAST the target and attribute pictures are very similar, participants may frequently fail to shift between tasks, leading to relatively many errors, causing the expected effects to be expressed most clearly in the ER data. Following this line of reasoning one would expect stronger ER EAST-effects on task-shift trials, in particular on target trials that elicit very strong evaluative associations. In the present study, task-shift effects should thus be particularly pronounced on the positive and negative target trials. To explore this possibility we calculated separate EAST-scores on the basis of

task-shift and non-shift target trials. The task-shift and non-shift EAST-scores for the ER and RT data are presented in Figure 1. A 5 Category (positive, neutral, negative, spider, white) x 2 Task-Shift (task-shift, non-shift) x 4 version ANOVA with repeated measures showed that the crucial Category x Task-Shift interaction was not significant for the RT data, $F(4, 38) < 1$. Meanwhile, the ER data did clearly support a task shift account; that is the Category x Task-Shift interaction was significant for the ER data, $F(4, 41) = 3.8, p < .05, \eta^2 = .27$. Subsequent t-tests showed that the ER EAST-scores for positive were significantly higher, and for negative significantly lower for task-shift than for non-shift trials, $t(47) = 2.9, p < .01$, and $t(47) = -2.4, p < .05$, respectively.

RT EAST-scores as a function of task-shift



ER EAST-scores as a function of task-shift

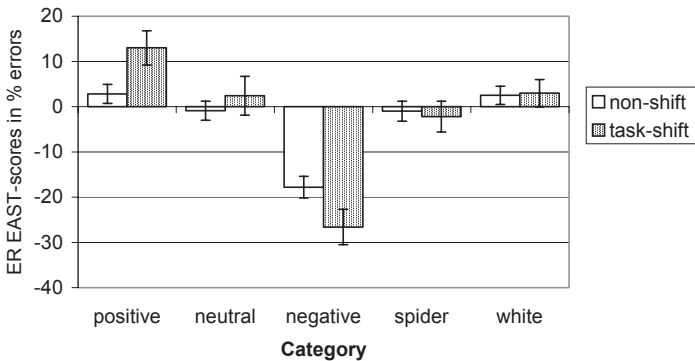


Figure 1.

Mean ER and RT based EAST-scores as a function of target category and task-shift.

A potential problem that follows from this account is that some participants may have sorted some or even all stimuli on the basis of their valence, because the task was simply too difficult. ER EAST-scores based on trials on which a participant failed to shift between tasks provide no reliable indication of *automatic* affective associations with the target stimuli. Yet, there are good arguments why participants probably did their best to sort the target pictures on the basis of their form, rather than on the basis of their valence. First, when an error was made error feedback was presented. Second, and probably more important, after an error the stimulus remained on the screen until the correct response was given. Thus, sorting all stimuli on the basis of their valence would be a rather inefficient strategy. Nevertheless, these findings suggest that one should be cautious with including individuals with many errors in the analysis, as their EAST-scores may not provide a valid index of automatic associations with the target stimuli. The fact that in the present sample 11 participants had an error rate over 30% suggests that although the task can be used in a community sample it may be too difficult for some individuals. It is important to note, however, that although the excluded participants were significantly older than the included participants, no significant correlation emerged between participants' age and the size of the EAST-scores, even when including all participants in the analysis.

A related issue is that there may be stable interindividual differences in the ability to shift between tasks. This would mean that interindividual differences in the size of the ER EAST-scores could at least partly reflect task-shift abilities rather than the strength of automatic associations. Indeed, there is evidence from research with the IAT that there are interindividual differences in individuals' ability to shift between tasks (e.g., Mierke & Klauer, 2001). However, the difference between the task-shift and non-shift ER EAST-scores for positive and negative pictures was not significantly related to participants' age (for positive, $r = .12$, $p > .4$; for negative, $r = -.24$, $p > .1$) or level of education (for positive, $r = -.13$, $p > .4$; for negative, $r = .11$, $p > .4$). This provides at least some indication that general task-shift ability had no major effect on the present data. Nevertheless, there may be other factors involved in individuals' general task-shift ability that are not directly related to age or educational level. It would therefore be important for future research to more formally tease apart the influence of task-shifting ability and associative strength on interindividual differences in EAST-scores.

The reliability of the EAST-scores ranged mostly from medium to quite reliable (see Table 1). This is somewhat surprising for the normative items as these items are not expected to yield stable interindividual differences. These findings are probably due to the diversity of the present sample and may, as argued above, reflect interindividual differences that are not directly related to the strength of affective associations.

Table 2.
 Correlations Between Age, Educational Level, the RT and ER based EAST-Scores for Positive, Neutral, and Negative Pictures, and the overall percentage of errors made while completing the EAST.

Measure	Age	RT			ER			Overall % errors
		Pos	Neu	Neg	Pos	Neu	Neg	
1. Age [†]	-	.01	-.28*	-.16	.10	.10	.00	.29*
2. Educational Level ^{††}	-.17	.00	-.01	-.13	-.02	-.19	-.05	-.16

Note. [†] n = 57; ^{††} n = 58; * p < .05.

Similar to the previous findings with the pictorial EAST (Huijding & de Jong, 2005), overall RTs were substantially slower and ERs more frequent than previously reported results with the verbal EAST (e.g., De Houwer, 2003). This could be due to the nature of the stimuli: The pictorial stimuli may have been more complex than words. Meanwhile, the present sample was also overall slower and less accurate than the student sample in the previous pictorial EAST study (Huijding & de Jong, 2005). Thus, the poorer overall task performance appears to have been caused by the fact that the present sample did not consist of graduate students. However, as can be seen in Table 2, there appeared no consistent relationships between participants' age or educational level and the EAST-scores. Only age and the RT-based EAST-score for neutral pictures showed a small but significant (negative) correlation, $r = -.28$, $p < .05$, suggesting that a higher age was associated with relatively more negative associations with the neutral pictures. Thus the present data suggest that even though non-student samples may show an overall somewhat poorer performance on the pictorial EAST, this does not affect the sensitivity of the EAST for affective associations. In addition, age and educational level appear not to be related to the size of the EAST-scores. However, as mentioned earlier, there may be other factors that influence the size of EAST-effects.

The present findings successfully replicated previous findings with the pictorial EAST, thereby underlining the reliability of the effects. The fact that the present findings are based on a community sample that varied considerably in age and educational level is encouraging for the width of possible applications of the task. However, the finding that EAST-effects are stronger for task-shift than for non-shift trials urges caution with respect to interpreting the EAST-scores of individuals with high error rates as reflecting automatic associations. A possible solution may be to simplify the task for instance by including category labels in the upper corners of the screen to remind participants of the task instructions. A limitation of the present study is that we used exclusively female sample that was not completely randomly recruited. Therefore, some form of selection bias cannot be precluded. In addition, it remains to be seen whether similar results would emerge for males. However, given the fact that we used universally positive and negative stimuli, it seems unlikely that a sample of male subjects would show a diverging pattern of results.

All in all, the present results underline that the modified EAST is sensitive to the affective value of pictorial stimuli independent of age and educational level. This finding opens the avenue for applying the EAST also in non-student samples and to concepts that are difficult to verbalise.

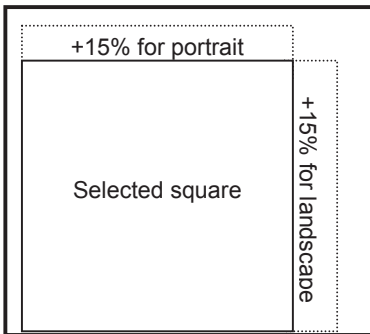
References

- Constantine, R., McNally, R.J., & Hornig, C.D. (2001). Snake fear and the pictorial Stroop paradigm. *Cognitive Therapy and Research, 25*, 757-764.
- De Houwer, J. (2003). The Extrinsic Affective Simon task. *Experimental Psychology, 50*, 77-85.
- De Houwer, J., Crombez, G., Koster, E.H.W., & De Beul, N. (2004). Implicit alcohol-related cognitions in a clinical sample of heavy drinkers. *Journal of Behavior Therapy and Experimental Psychiatry, 35*, 275-286.
- De Houwer, J., & d'Ydewalle, G. (1994). Stroop-like interference in sorting for intrinsic color: a test of the Glaser and Glaser (1989) model. *Acta Psychologica, 85*, 123-137.
- 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 Jong, P.J., van den Hout, M.A., Rietbroek, H., & Huijding, J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion, 17*, 521-545.
- de Jong, P.J., Wiers, R.W., & van de Braak, M. & Huijding, J. (in press). Using the Extrinsic Affective Simon Task as a measure of implicit attitudes towards alcohol: relationship with drinking behavior and alcohol problems. *Addictive Behaviors*.
- 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 Behavior 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.
- Glaser, W.R., & Glaser, W.O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*, 13-42.
- 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. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*, 85-93.
- Huijding, J., & de Jong, P.J. (2005). A pictorial version of the Extrinsic Affective Simon Task: sensitivity to generally affective and phobia-relevant stimuli in high and low spider fearful individuals. *Experimental Psychology, 52*, 289-295.
- Huijding, J., & de Jong, P.J. (2006). Specific predictive power of automatic spider-related affective associations for controllable and uncontrollable fear responses toward spiders. *Behaviour Research and Therapy, 44*, 161-176.
- Huijding, J., de Jong, P.J., Wiers, R.W., & Verkooijen, K. (2005). Implicit and explicit attitudes toward smoking in a smoking and a nonsmoking setting. *Addictive Behaviors, 30*, 949-961.
- Kindt, M., & Brasschot, J.F. (1999). Cognitive bias in spider-phobic children: comparison of a pictorial and a linguistic spider Stroop. *Journal of Psychopathology and Behavioral Assessment, 21*, 207-220.

- Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (1996). *The international affective picture system*. Gainesville, FL: NIHMH centre for the study of emotion and attention, University of Florida.
- Lavy, E., & van den Hout, M. (1993). Selective attention evidenced by pictorial and linguistic Stroop tasks. *Behavior Therapy*, 24, 645-657.
- 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.
- Schneider, W. (1989). Enhancing a standard experimental delivery system (MEL) for advanced psychological experimentation. *Behavior Research Methods, Instruments and computers*, 21, 240-244.
- Vansteenwegen, D., Crombez, G. Baeyens, F., & Eelen, P. (1998). Extinction in fear conditioning: effects on startle modulation and evaluative self-reports. *Psychophysiology*, 35, 729-736.
- Voss, A., Schmitz, F, Teige, S., & Klauer, C. (2005 april). *Cognitive capacity, task shifts, and measurement issues in the EAST procedure*. Data presented at the EAST workshop in Jena, Germany.

Appendix

A 'portrait' and a 'landscape' variant of each oblong picture was created in Photoshop by first selecting a square containing the most important part of the original for the target picture and then adding an additional section of approximately 15% of the total length of one side of the selected square to the top or the side of the selection (See below). All square pictures were made by placing a square selection of the original for the attribute picture upon a slightly larger yellow rectangle (grey in the figure below), resulting in a picture with a yellow border of 8 pixels wide. Finally, all pictures were standardised qua size. The square pictures were resized to 400 x 400 pixels (yellow border included). All oblong pictures were resized to 5 dimensions, the longest sides measuring 360, 380, 400, 420, or 440 pixels.

Original for target picture**Original for attribute picture**