

**A Thousand Words are Worth More Than a Picture?**

**The Effects of Stimulus Modality on the Implicit Association Test**

Franziska Meissner

Klaus Rothermund

**Online Supplement**

## Experiment 1

### ReAL Model Parameters

As reported in the main text, the ReAL model (Meissner & Rothermund, 2013) not only provides estimates for recoding (*Re*) and evaluative associations (*A*), but also for a label-based identification of the correct response (*L*), as well as estimates for three technical parameters reflecting the asymmetries in the cognitive processes with respect to task switch vs. task repetition trial sequences. We report the mean parameter estimates for these additional model parameters in Table S1.

### Method

**Materials.** We ran a pilot study in order to ensure that our pictorial and verbal materials would not differ systematically in valence or category typicality: Twenty participants evaluated a list of verbal and pictorial flowers, insects, positive and negative stimuli (7-point scale; higher values indicate more positive evaluations). Furthermore, they rated the category typicality of the flower and insect stimuli (7-point scale; higher values indicate higher typicality for the particular category). Based on those ratings, we chose the stimulus materials for the current experiment so that the verbal and pictorial sets were comparable in terms of valence and typicality, for the category *flower* (words:  $M_{\text{valence}} = 5.33$ ,  $SD = 1.06$ ;  $M_{\text{typicality}} = 5.70$ ,  $SD = 1.07$ ; pictures:  $M_{\text{valence}} = 5.59$ ,  $SD = 1.05$ ;  $M_{\text{typicality}} = 5.65$ ,  $SD = 1.33$ ), *insect* (words:  $M_{\text{valence}} = 2.19$ ,  $SD = 0.98$ ;  $M_{\text{typicality}} = 5.33$ ,  $SD = 1.07$ ; pictures:  $M_{\text{valence}} = 2.09$ ,  $SD = 0.88$ ;  $M_{\text{typicality}} = 5.41$ ,  $SD = 1.33$ ), *positive* (words:  $M_{\text{valence}} = 6.39$ ,  $SD = 0.66$ ; pictures:  $M_{\text{valence}} = 6.41$ ,  $SD = 0.61$ ), and *negative* (words:  $M_{\text{valence}} = 1.66$ ,  $SD = 0.54$ ; pictures:  $M_{\text{valence}} = 1.65$ ,  $SD = 0.70$ ). For each of the four categories, *flower*, *insect*, *positive* and *negative*, we selected four verbal exemplars (e.g., *orchid*, *cockroach*, *peace*, *hostility*) as well as four pictorial stimuli (e.g., colored photographs of a carnation, a fly, a sunset, an explosion). In the IAT, pictorial stimuli were displayed on the screen in landscape format with

a size of about 9.5 x 7.2 cm, and all verbal stimuli were printed in black font on a white background.

**Table S1**

Mean estimates for the additional ReAL model parameters in Experiments 1 and 2 (standard errors in parentheses)

	Verbal Attributes		Pictorial Attributes	
	Verbal Targets	Pictorial Targets	Verbal Targets	Pictorial Targets
<b>Experiment 1</b>				
<i>L<sub>flower</sub></i>	.54 (.04)	.71 (.04)	.44 (.04)	.66 (.03)
<i>L<sub>insect</sub></i>	.50 (.02)	.67 (.03)	.44 (.05)	.65 (.03)
<i>L<sub>positive</sub></i>	.72 (.05)	.59 (.05)	.75 (.03)	.72 (.05)
<i>L<sub>negative</sub></i>	.68 (.04)	.52 (.04)	.78 (.04)	.75 (.03)
<i>attL</i>	.45 (.07)	.43 (.05)	.78 (.04)	.55 (.05)
<i>attReC</i>	.44 (.10)	.50 (.11)	.27 (.09)	.30 (.09)
<i>attReT</i>	.44 (.10)	.26 (.09)	.31 (.08)	.31 (.09)
<b>Experiment 2</b>				
<i>L<sub>young</sub></i>	.78 (.03)	.77 (.03)	.69 (.04)	.70 (.03)
<i>L<sub>old</sub></i>	.64 (.03)	.73 (.03)	.63 (.05)	.73 (.04)
<i>L<sub>positive</sub></i>	.77 (.04)	.75 (.05)	.78 (.03)	.72 (.04)
<i>L<sub>negative</sub></i>	.72 (.05)	.75 (.04)	.82 (.03)	.82 (.03)
<i>attL</i>	.43 (.06)	.68 (.05)	.63 (.07)	.58 (.06)
<i>attReC</i>	.45 (.09)	.64 (.10)	.35 (.09)	.58 (.10)
<i>attReT</i>	.37 (.08)	.23 (.08)	.20 (.09)	.27 (.08)

**Note.** *L* = label-based identification of the correct response (estimated separately for each stimulus category). *attL* = attenuation of *L* for task switch sequences compared to task repetition sequences. *attReC* = attenuation of *Re* for the target categories compared to the attribute categories. *attReT* = attenuation of *Re* in task repetition trials compared to task switch trials.

**Procedure.** The IAT followed the procedure suggested by Meissner and Rothermund (2013). More precisely, participants were required to speed up their responses, and to tolerate a higher error rate. Additionally, an individual response deadline procedure was employed and an extra reward was provided if participants met this response speed criterion sufficiently often. Following Meissner and Rothermund (2013), we also augmented the number of IAT trials, and split the single pair of compatible and incompatible block into a number of short IAT block pairs which had to be performed in succession.

Participants started with a simple attribute categorization practice block (16 trials) followed by a target categorization practice block (16 trials). Afterwards, participants worked through the first combined block where stimuli of both dimensions had to be classified. Following that, the target classification was practiced with a reversed response assignment (16 trials). Then, the second combined block was presented, which was identical to the first, except for the reversed target assignment. This initial pair of combined blocks served as warm-up blocks and, most important, as data base for a calibration of the individual response deadline. Afterwards, participants performed five further pairs of combined blocks. The block order within the block pairs (compatible first vs. incompatible first) was counterbalanced and did not interact with the experimental factors. Each combined block comprised 32 trials (plus 1 dummy trial in the beginning). Stimuli were presented in a randomized order with the exception that all kinds of task switch sequences occurred equally often for each stimulus category. In order to control for individual differences and practice effects, the individual response deadline was adapted after each block pair based on the error level in the previous block pair (see Meissner & Rothermund, 2013).

In all blocks, the category labels were shown in the upper right and upper left part of the screen, and participants had to assign the stimuli to the correct category by pressing one of the two response keys (i.e., “D” or “L” on the computer keyboard). On each trial, a stimulus was presented in the center of the screen until the correct response key was pressed. If

participants did not respond before the individual deadline, the stimulus was framed by a red rectangle in order to remind participants to respond more quickly. In case of an incorrect response, an error message was presented until the correct response was given. The intertrial interval was 200 ms.

In preparation of data analysis, we again followed Meissner and Rothermund (2013) and excluded all practice trials, the first block pair, as well as the first trial of all combined test blocks from analysis.

## Results

Additionally to error based IAT scores, we also computed scores based on the widely used *D* scoring algorithm (with built-in error penalty; Greenwald, Nosek, & Banaji, 2003; see Table S2). A 2 (target modality: words vs. pictures) x 2 (attribute modality: words vs. pictures) ANOVA of these *D* scores also revealed a significant interaction,  $F(1, 76) = 12.42, p < .001, \eta_p^2 = .14$  (main effects were *ns*, all  $F_s \leq 1.29, p \geq .259$ ). The simple effects were in the expected directions,  $t(76) = 1.69, p = .095, d = 0.53$ , and,  $t(76) = -3.29, p = .001, d = -1.04$ .

**Table S2**

Mean *D* scores in the four groups for Experiment 1 and 2 (standard errors in parentheses)

	Verbal Attributes		Pictorial Attributes	
	Verbal Targets	Pictorial Targets	Verbal Targets	Pictorial Targets
Experiment 1	.45 (5.07)	.36 (4.13)	.31 (3.04)	.50 (3.44)
Experiment 2	.54 (4.80)	.33 (3.84)	.34 (7.99)	.37 (6.28)

*Note.* *D* Score = the improved scoring algorithm for the IAT (Greenwald et al., 2003). Larger scores were interpreted in terms of a larger preference of flowers over insects and of young over old age, respectively.

## Discussion

In order to bolster our conclusions based on the pattern of results in Experiment 1, we collected additional data from a *standard* flower-insect IAT ( $N = 64$ ) with the same pre-tested

stimulus material. In this replication study, attribute modality was manipulated between-subjects while target modality was a within-subjects factor: Each participant completed two IATs that differed in the modality of the target stimuli. The order of IATs was counterbalanced. The traditional 7-steps IAT procedure (attribute classification, target classification, first combined practice block, first combined test block, reversed target classification, second combined practice block, second combined test block) was employed with counterbalanced block order. We included standard IAT instructions (“Please try to respond as fast and as accurately as possible.”) and omitted the response deadline. Consequently, the focus of this replication study was not on error rates but on response time effects. As expected, both the *D* score and the standard response time score revealed significant modality match effects as indicated by a significant interaction of target and attribute modality,  $F(1, 62) = 4.25, p = .043, \eta_p^2 = .06$ , and  $F(1, 62) = 4.35, p = .041, \eta_p^2 = .07$  (see Table S3). Both simple effects revealed the expected directions, for *D* scores,  $t(62) = 1.08, p = .285, d = 0.19$ , and,  $t(62) = -1.84, p = .071, d = -0.32$ , as well as response time scores,  $t(62) = 1.14, p = .258, d = 0.20$ , and,  $t(62) = -1.81, p = .075, d = -0.32$ . We were thus successful in replicating the results of Experiment 1 in a standard IAT procedure.

**Table S3**

Mean *D* scores and response times in the replication study (standard errors in parentheses)

	Verbal Attributes		Pictorial Attributes	
	Verbal Targets	Pictorial Targets	Verbal Targets	Pictorial Targets
<i>D</i> Score	.71 (.06)	.63 (.06)	.56 (.07)	.70 (.05)
RT Score	178 (19)	158 (19)	113 (14)	146 (18)

*Note.* *D* Score = the improved scoring algorithm for the IAT (Greenwald et al., 2003). RT Score = difference in mean response times (in ms) between compatible and incompatible blocks. Larger scores were interpreted in terms of a larger preference of flowers over insects and of young over old age, respectively.

## Experiment 2

### Participants

Three participants were excluded from analysis: One of them reported that he was not able to classify the names unambiguously as old or young. One participant reported that he did not understand the IAT test block instructions after completing the experiment. Finally, one participant produced a mean error rate (41%) that was a far outlier with respect to the sample distribution (deviating more than 4.5 standard deviations from the sample mean).

### Method

**Materials.** The material was chosen based on a pilot study with a separate sample: Sixteen students evaluated several old and young names (chosen from the most popular forenames for children born in Germany in 1945 and 1995, respectively) and faces (from the FACES database; Ebner, Riediger, & Lindenberger, 2010) as well as several positive and negative words and pictures (7-point scale; higher values indicate more positive evaluations). Furthermore, they rated the age typicality of the stimuli (7-point scale; higher [lower] values indicate a high typicality for old [young] age). The stimuli were chosen so that the verbal and pictorial sets were comparable in terms of valence and age typicality, for the category *young* (words:  $M_{\text{valence}} = 4.39$ ,  $SD = 0.74$ ;  $M_{\text{age}} = 2.55$ ,  $SD = 0.50$ ; pictures:  $M_{\text{valence}} = 4.53$ ,  $SD = 0.71$ ;  $M_{\text{age}} = 2.69$ ,  $SD = 0.53$ ), *old* (words:  $M_{\text{valence}} = 3.45$ ,  $SD = 0.97$ ;  $M_{\text{age}} = 5.66$ ,  $SD = 0.74$ ; pictures:  $M_{\text{valence}} = 3.69$ ,  $SD = 0.63$ ;  $M_{\text{age}} = 5.80$ ,  $SD = 0.46$ ), *positive* (words:  $M_{\text{valence}} = 6.38$ ,  $SD = 0.54$ ;  $M_{\text{age}} = 3.89$ ,  $SD = 0.96$ ; pictures:  $M_{\text{valence}} = 6.41$ ,  $SD = 0.46$ ;  $M_{\text{age}} = 3.86$ ,  $SD = 0.67$ ), and *negative* (words:  $M_{\text{valence}} = 1.42$ ,  $SD = 0.51$ ;  $M_{\text{age}} = 4.23$ ,  $SD = 0.87$ ; pictures:  $M_{\text{valence}} = 1.48$ ,  $SD = 0.61$ ;  $M_{\text{age}} = 4.03$ ,  $SD = 0.34$ ). For each of the four categories, *young*, *old*, *positive* and *negative*, we selected four verbal exemplars (e.g., *Tobias*, *Helmut*, *vacation*, *violence*) as well as four pictorial stimuli (e.g., a face of a young man, a face of an old man, a beach, a shark). Half of the faces as well as half of the names were female. Each of the face

stimuli portrayed a white person with a neutral expression wearing a grey t-shirt on a dark background. All pictorial stimuli were colored photographs in portrait format, and all verbal stimuli were printed in black font on a white background.

**Procedure.** The procedure of this age attitude IAT was identical to Experiment 1 with the exception that the response deadline was now fixed to 700 ms for all participants and for all IAT block pairs. We included this procedural modification in order to prevent a confound of the experimental between-subjects factors and features of the response deadline. Hence, any effects that were obtained in this experiment cannot be attributed to differences in the procedural setting. By applying a fixed response deadline, however, we allowed for more noise in our data, since the criterion would be easier for some participants than for others. This might result in very low and very high error rates, respectively, and therefore affects the reliability of the individual parameter estimates.

Analysis was prepared as in Experiment 1 with the exception that we refrained from excluding the first block pair from the analyses (due to the omission of the individual deadline adaptation, the first block pair did not differ procedurally from the subsequent blocks).

## Results

A 2 (target modality: words vs. pictures) x 2 (attribute modality: words vs. pictures) ANOVA with the *D* Score as the dependent variable revealed a marginally significant interaction of target modality x attribute modality,  $F(1, 73) = 3.84, p = .054, \eta_p^2 = .05$  (all main effects were *ns*, all  $F_s \leq 2.09, p \geq .152$ ). Simple effects tests confirmed significantly smaller IAT scores for pictorial targets than for target words,  $t(73) = 2.39, p = .019, d = 0.78$ , if the attribute stimuli were *words*. However, if *pictures* were used as attributes, this effect of target modality vanished,  $t(73) = -0.36, p = .716, d = -0.12$  (see Table S1).

## References

- Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES – A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. doi:10.3758/BRM.42.1.351
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. doi:10.1037/0022-3514.85.2.197
- Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45–69. doi:10.1037/a0030734