Supplemental Online Material

of the Article

Estimating the Contributions of Associations and Recoding in the Implicit Association Test:

The ReAL Model for the IAT

Franziska Meissner and Klaus Rothermund

Friedrich-Schiller-University Jena, Germany

ReAL Model Equations for the Flower-Insect IAT

For the compatible assignment:

$p(+ \mid \text{flower, tr}) =$ $Re*attReC*attReT + Re*attReC*(1-attReT)*L_1 + Re*attReC*(1-attReT)*(1-L_1)*A_1 + Re*(1-attReC)*L_1 + Re*(1-attReC)*(1-L_1)*A_1 + (1-Re)*L_1 + (1-Re)*(1-L_1)*A_1$

$p(- \mid \text{flower, tr}) =$ $Re*attReC*(1-attReT)*(1-L_1)*(1-A_1) + Re*(1-attReC)*(1-L_1)*(1-A_1) + (1-Re)*(1-L_1)*(1-A_1)$

$p(+ \mid \text{flower, ts}) =$ $Re*attReC + Re*(1-attReC)*L_1*attL + Re*(1-attReC)*L_1*(1-attL)*A_1 + Re*(1-attReC)*(1-L_1)*A_1 + (1-Re)*L_1*attL + (1-Re)*L_1*(1-attL)*A_1 + (1-Re)*(1-L_1)*A_1$

$p(- \mid \text{flower, ts}) =$ $Re*(1-attReC)*L_1*(1-attL)*(1-A_1) + Re*(1-attReC)*(1-L_1)*(1-A_1) + (1-Re)*L_1*(1-attL)*(1-A_1) + (1-Re)*(1-L_1)*(1-A_1)$

$p(+ \mid \text{insect, tr}) =$ $Re*attReC*attReT + Re*attReC*(1-attReT)*L_2 + Re*attReC*(1-attReT)*(1-L_2)*(1-A_2) + Re*(1-attReC)*L_2 + Re*(1-attReC)*(1-L_2)*(1-A_2) + (1-Re)*L_2 + (1-Re)*(1-L_2)*(1-A_2)$

$p(- \mid \text{insect, tr}) =$ $Re*attReC*(1-attReT)*(1-L_2)*A_2 + Re*(1-attReC)*(1-L_2)*A_2 + (1-Re)*(1-L_2)*A_2$

$p(+ \mid \text{insect, ts}) =$ $Re*attReC + Re*(1-attReC)*L_2*attL + Re*(1-attReC)*L_2*(1-attL)*(1-A_2) + Re*(1-attReC)*(1-L_2)*(1-A_2) + (1-Re)*L_2*attL + (1-Re)*L_2*(1-attL)*(1-A_2) + (1-Re)*(1-L_2)*(1-A_2)$

$p(-\mid \text{insect, ts}) =$ $Re*(1\text{-}attReC)*L_2*(1\text{-}attL)*A_2 + Re*(1\text{-}attReC)*(1\text{-}L_2)*A_2 + (1\text{-}$

  $Re)*L_2*(1\text{-}attL)*A_2 + (1\text{-}Re)*(1\text{-}L_2)*A_2$

$p(+\mid \text{good, tr}) =$ $Re*attReT + Re*(1\text{-}attReT)*L_3 + Re*(1\text{-}attReT)*(1\text{-}L_3)*.5 + (1\text{-}Re)*L_3 +$

  $(1\text{-}Re)*(1\text{-}L_3)*.5$

$p(-\mid \text{good, tr}) =$ $Re*(1\text{-}attReT)*(1\text{-}L_3)*(1\text{-}.5) + (1\text{-}Re)*(1\text{-}L_3)*(1\text{-}.5)$

$p(+\mid \text{good, ts}) =$ $Re + (1\text{-}Re)*L_3*attL + (1\text{-}Re)*L_3*(1\text{-}attL)*.5 + (1\text{-}Re)*(1\text{-}L_3)*.5$

$p(-\mid \text{good, ts}) =$ $(1\text{-}Re)*L_3*(1\text{-}attL)*(1\text{-}.5) + (1\text{-}Re)*(1\text{-}L_3)*(1\text{-}.5)$

$p(+\mid \text{bad, tr}) =$ $Re*attReT + Re*(1\text{-}attReT)*L_4 + Re*(1\text{-}attReT)*(1\text{-}L_4)*(1\text{-}.5) + (1\text{-}Re)*L_4$

  $+ (1\text{-}Re)*(1\text{-}L_4)*(1\text{-}.5)$

$p(-\mid \text{bad, tr}) =$ $Re*(1\text{-}attReT)*(1\text{-}L_4)*.5 + (1\text{-}Re)*(1\text{-}L_4)*.5$

$p(+\mid \text{bad, ts}) =$ $Re + (1\text{-}Re)*L_4*attL + (1\text{-}Re)*L_4*(1\text{-}attL)*(1\text{-}.5) + (1\text{-}Re)*(1\text{-}L_4)*(1\text{-}.5)$

$p(-\mid \text{bad, ts}) =$ $(1\text{-}Re)*L_4*(1\text{-}attL)*.5 + (1\text{-}Re)*(1\text{-}L_4)*.5$

For the incompatible assignment:

$p(+\mid \text{flower, tr}) =$ $L_1 + (1\text{-}L_1)*(1\text{-}A_1)$

$p(-\mid \text{flower, tr}) =$ $(1\text{-}L_1)*A_1$

$p(+\mid \text{flower, ts}) =$ $L_1*attL + L_1*(1\text{-}attL)*(1\text{-}A_1) + (1\text{-}L_1)*(1\text{-}A_1)$

$p(-\mid \text{flower, ts}) =$ $L_1*(1\text{-}attL)*A_1 + (1\text{-}L_1)*A_1$

$p(+ \mid \text{insect, tr}) =$ $\quad L_2 + (1-L_2)*A_2$

$p(- \mid \text{insect, tr}) =$ $\quad (1-L_2)*(1-A_2)$

$p(+ \mid \text{insect, ts}) =$ $\quad L_2*attL + L_2*(1-attL)*A_2 + (1-L_2)*A_2$

$p(- \mid \text{insect, ts}) =$ $\quad L_2*(1-attL)*(1-A_2) + (1-L_2)*(1-A_2)$

$p(+ \mid \text{good, tr}) =$ $\quad L_3 + (1-L_3)*(1-.5)$

$p(- \mid \text{good, tr}) =$ $\quad (1-L_3)*.5$

$p(+ \mid \text{good, ts}) =$ $\quad L_3*attL + L_3*(1-attL)*(1-.5) + (1-L_3)*(1-.5)$

$p(- \mid \text{good, ts}) =$ $\quad L_3*(1-attL)*.5 + (1-L_3)*.5$

$p(+ \mid \text{bad, tr}) =$ $\quad L_4 + (1-L_4)*.5$

$p(- \mid \text{bad, tr}) =$ $\quad (1-L_4)*(1-.5)$

$p(+ \mid \text{bad, ts}) =$ $\quad L_4*attL + L_4*(1-attL)*.5 + (1-L_4)*.5$

$p(- \mid \text{bad, ts}) =$ $\quad L_4*(1-attL)*(1-.5) + (1-L_4)*(1-.5)$

$+$ = correct response; $-$ = incorrect response; tr = task repetition; ts = task switch. Re = activation of the recoded response category; A = evaluative associations; L = label-based identification of the correct response; $attL$ = attenuation of L for task switch trials; $attReT$ = attenuation of Re for task repetition trials; $attReC$ = attenuation of Re for the target categories. The value .5 in the model equations reflects the restriction of the association parameters for the attribute categories to the neutral point: $A_3 = A_4 = .5$.

Technical Details of the ReAL Model

Database and model fit. The database of the ReAL model consists of correct and incorrect responses in each of the four different stimulus categories within the compatible and the incompatible block, further separated into task repetition and task switch trial sequences. In sum, we can observe 16 non-redundant response categories per IAT for each participant. The ReAL model explains these observable response categories with seven parameters (one *Re*, two *A*, and four *L* parameters). Adding three technical parameters (see below), we can test the model's fit to IAT data with six degrees of freedom (i.e., 16 non-redundant response categories - 10 model parameters = 6 degrees of freedom).

Note that in the literature concerning multinomial models, parameter estimation was often based on aggregated data (i.e., response frequencies that were summed up across participants; e.g., Conrey et al., 2005; Payne et al., 2010; Stahl & Degner, 2007). However, as the IAT was developed and often applied as a measure of interindividual differences in attitudes, an aggregation across participants does obviously not provide an adequate analytic strategy. There are two alternative approaches which can be used instead: First, the latent-class approach by Klauer (2006) can be employed by dividing the sample into several latent classes via hierarchical multinomial models. Different parameter values can be estimated for each latent class so that different predictions for several groups of participants are possible (see also Klauer, 2010, for a different approach based on hierarchical models). Second, parameters can simply be estimated based on the individual rather than the aggregated response frequencies. We decided to use the latter approach and adapted the IAT procedure in order to obtain an optimal data base for individual parameter estimation.

Based on the observed response pattern and the expectation-maximization mechanism (e.g. Batchelder & Riefer, 1999), the ReAL model parameters can be estimated so that a minimum of the log-likelihood statistic G^2 is obtained. This G^2 statistic which is approximately chi-square distributed for large samples represents the divergence of the observed response frequencies from the pattern that can be expected based on the ReAL model equations. If the G^2 statistic is non-significant, the assumption of a fit to the data can be held.

Recoding in the incompatible block. Two anonymous reviewers suggested that recoding could also play some role in the incompatible block. If it is assumed that the recoded category from the compatible block at least sometimes becomes activated in the incompatible block as well, recoding could lead to incorrect responses in the incompatible block. On the other hand, if a different recoded category is used in the incompatible block in comparison to the compatible block (e.g. because of ambivalent targets), recoding would lead to the correct response in this block. Our model allows for testing these assumptions by estimating a separate recoding parameter for the incompatible block. It turned out that this new parameter is almost never significantly different from zero, whether it is modeled to produce incorrect or correct responses: Across all experiments, only 3 out of 402 participants had a significant loss of model fit if recoding was restricted to zero in the incompatible block (applying the Bonferroni-Holm correction). Thus, recoding did not play a meaningful role in the incompatible block. We decided to keep the model as simple as possible and included the recoding parameter only in the compatible block (restricting it to zero in the incompatible block).

Technical attenuation parameters. In order to increase the model's fit, three technical parameters were included in the ReAL model. These parameters, mainly reflecting the asymmetry of parameters between task switch and task repetition trial sequences, are only of

technical relevance and are therefore not mentioned in the experimental results. Furthermore, the individual confidence intervals for these technical parameters are rather large, and often cover the whole parameter range (i.e., they are not different from 0 and from 1 at the same time). Interpreting and testing these individual parameter estimates is thus difficult. However, as a fixation of these parameters revealed a significant loss of model fit for several applications, we decided to include these parameters permanently in the ReAL model. They were estimated in order to map the corresponding attenuation so that it does not bias the other model parameters. Thus, through mapping the mentioned differences between task switch and task repetition trials on the technical parameters, we could increase the model's fit and the validity of the relevant parameters. If, for example, the fluctuating categorization difficulty between task switch and task repetition trials was ignored, the model fit would be almost certainly harmed in most applications. The logic behind these technical parameters is introduced in the following paragraph.

Given that task switches involve costs in task performance, the controlled label-based identification process should be more difficult after a task switch compared to a task repetition. In order to implement this assumption in the model equations, we installed a technical parameter reflecting an order constraint (Knapp & Batchelder, 2004). Such order restrictions are reparametrizations of the model which secured that a parameter could be slightly smaller in one condition compared to the other. Thus, beside the four *L* parameters, an additional attenuation parameter (*attL*) is estimated reflecting the attenuation of *L* for task switch sequences compared to task repetition sequences. If this technical parameter was, for example, .75 then we would conclude that *L* is a quarter smaller in task switch compared to task repetition trials. Furthermore, we included the technical parameter *attReT* referring to the attenuation of *Re* in task repetition

trials compared to task switch trials: Due to the reduced difficulty of the categorization task in repetition sequences, simply the same response set could be retrieved so that the probability of activating the recoded response category is at most equal but probably smaller than in task switch trials. Finally, the technical parameter *attReC* was included, basically reflecting an attenuation of *Re* between the categories: In most attitude IATs, targets (e.g., *flower* and *insect*) are recoded in terms of their attribute characteristics (i.e., *good* and *bad*). It seems plausible that *flower* and *insect* would activate the recoded response category to a slightly less extent than the attributes. Thus, beside the common *Re* parameter, an additional attenuation parameter (*attReC*) is estimated for the target categories reflecting this attenuation. Such a parameter, however, could not be identified unless the response set is split into task repetition and task switch trials. The conducted split of the database thus not only increases the model's fit, it also provides enough response categories for identifying all model parameters. Finally, it allows for reasonable restrictions of parameters (e.g., the attenuation of label-based identification in task switch trials is equal for all stimulus categories) and thus increases the face validity of the model parameters. The mean parameter estimates for the three attenuation parameters are presented in Supplemental Table 1.

Supplemental Table 1

Mean parameter estimates for attenuation parameters for all experiments (standard errors in parentheses)

Experiment	attenuation of L for task switch ($attL$)	attenuation of Re for task repetition ($attReT$)	attenuation of Re for two categories ($attReC$)
Experiment 1	.51 (.05)	.28 (.06)	.29 (.06)
Experiment 2	.49 (.03)	.27 (.06)	.53 (.07)
Experiment 3	.77 (.02)	.38 (.05)	.37 (.05)
Experiment 4	.62 (.04)	.34 (.07)	.27 (.06)
Experiment 5	.49 (.04)	.35 (.07)	.61 (.07)
Experiment 6			
Females	.43 (.04)	.30 (.06)	.87 (.04)
Males	.64 (.03)	.18 (.06)	.67 (.07)
Experiment 7	.60 (.02)	.17 (.04)	.37 (.05)

Analyses With the Quad Model

We used the currently recommended Quad model with two AC, one D, one G and one OB parameter. The corresponding equations differ to a certain extent from that reported by Conrey, Sherman, Gawronski, Hugenberg, and Groom (2005) and are available at http://psychology.ucdavis.edu/labs/sherman/site/research.html.

With the exception of Experiment 6 where a good fit was observed, median $G^2(3) = 4.47$, $p = .215$, the model fit never exceeded the significance level of $p = .10$ and was indeed significant for about half of the reported experiments, median $G^2(3) \geq 9.76$, $p \leq .021$. However, the Quad model fit statistic was only marginally significant and thus at least satisfying for Experiment 2, median $G^2(3) = 6.40$, $p = .094$, Experiment 5, median $G^2(3) = 6.92$, $p = .075$, and Experiment 7, median $G^2(3) = 7.55$, $p = .056$. The unsatisfactory model fit for many of the reported experiments was also obtained if only task switch trials are included in the model analysis, and it also holds if a less restrictive Quad model version with separate D parameters for attributes and targets is used.

As described above, we assumed that the Quad model confounds association and recoding processes in the AC parameter. We tested this hypothesis in Experiment 6 where the Quad model fitted the best. In two separate analyses, the two AC parameters of the Quad model were regressed on the corresponding *A* parameter as well as the *Re* parameter of the ReAL model. For reasons of clarity, we included only participants with a positive switch cost effect in this analyses although the conclusions were identical if the complete sample was taken into account. It turned out that in both regression analyses, not only the corresponding *A* parameter (in absolute values: both $\beta \geq .20$,), $t(59) \geq 2.05$, $p \leq .045$, but also the *Re* parameter were significant predictors of AC (both $\beta \geq .62$,), $t(59) \geq 6.43$, $p < .001$ ($R^2 \geq .46$). As expected, the

AC parameters of the Quad model seem to measure a mixture of recoding processes and evaluative associations.

Stimuli in Experiment 1 to 7

Flower (Experiment 1, 2, 3)

Flieder [lilac], Krokus [crocus], Lilie [lily], Nelke [carnation], Orchidee [orchid], Rose [rose], Tulpe [tulip], Veilchen [violet]

Insect (Experiment 1, 2, 3)

Ameise [ant], Floh [flea], Grille [cricket], Hornisse [hornet], Käfer [beetle], Made [maggot], Stechmücke [mosquito], Wespe [wasp]

Good (Experiment 1 to 5)

EHRLICH [honest][a], FRIEDEN [peace], GESUND [healthy][b], HUMOR [humor], LIEBE [love], SANFT [gentle], SOMMER [summer], TREU [faithful], URLAUB [vacation]

Bad (Experiment 1 to 5)

ABGAS [exhaust], ANGST [anxiety][b], BOMBE [bomb][a], EINSAM [lonely], ELEND [misery], GEIZIG [miserly][a], GEWALT [violence][b], GIFT [poison][a], GRAUSAM [cruel][b], KRIEG [war], SCHMERZ [pain][b], VERLUST [loss][a]

Soccer team 1 (Experiment 4)

Florian, Jens, Markus, Robert

Soccer team 2 (Experiment 4)

Andreas, Dirk, Stefan, Tobias

German (Experiment 5)

Anna, Daniel, Frank, Marie, Moritz, Susi, Thomas, Ute

Turkish (Experiment 5)

Ali, Ayse, Fatma, Hakan, Kiraz, Mehmet, Murat, Özlem

Women (Experiment 6)

Frau [woman], sie [she], Mädchen [girl], Dame [lady]

Men (Experiment 6)

Mann [man], er [he], Junge [boy], Herr [gentleman]

Good (Experiment 6)

FABELHAFT [fabulous], PERFEKT [perfect], POSITIV [positive], SUPER [super]

Bad (Experiment 6)

ELEND [misery], NEGATIV [negative], SCHLIMM [severe], SCHRECKLICH
[horrible]

Positive (Experiment 7)

ERFOLG [success], FRIEDEN [peace], HUMOR [humor], URLAUB [vacation]

Negative (Experiment 7)

ANGST [anxiety], GEWALT [violence], KRIEG [war], SCHMERZ [pain]

---

[a]Experiment 4 only. [b]Deleted in Experiment 4.