Supplementary online material for *Of Two Minds: A registered replication*

Tobias Heycke[1, 2], Frederik Aust[1], Mahzarin R. Banaji[3], John G. Conway[4], Pieter Van Dessel[5], Xiaoqing Hu[6], Congjiao Jiang[4], Benedek Kurdi[3], Robert Rydell[7], Lisa Spitzer[1], Christoph Stahl[1], Christine A. Vitiello[4], & Jan De Houwer[5]

[1] University of Cologne
[2] GESIS - Leibniz Institute for the Social Sciences
[3] Harvard University
[4] University of Florida
[5] Ghent University
[6] The University of Hong Kong
[7] Indiana University

Correspondence concerning this article should be addressed to Tobias Heycke, P.O. 122155, 68072 Mannheim, Germany. E-mail: tobias.heycke@gesis.org

Experiment 1

In the following we report additional analyses and provide details for the model specification used for the Bayesian model comparisons. We report results from the linear mixed model analysis of the IAT response times, from prior sensitivity analyses for the Bayesian model comparisons, and from an exploratory analysis of the relationship between US recognition accuracy and associative learning. Table 1 summarizes the participants' demographics separately for each location of data collection.

Table 1

Participant demographics by location.

Location	Age	Female (%)	*n*
Cologne	24.61 [18, 64]	70.59	51
Ghent	21.92 [17, 50]	82.00	50
Harvard	19.58 [18, 22]	57.69	52

Note. Mean age is given with range in brackets.

Mixed model analysis

The ANOVA of IAT scores reported in the main text ignores potential systematic trial-to-trial variability in IAT response latencies due to stimuli. Any such systematic but unaccounted-for variance can inflate test statistics and yield underestimated *p* values as well as underestimated confidence intervals. We, therefore, also conducted a linear mixed model analysis of response times with crossed random effects for participants and items to ensure that our conclusion are not contingent on inadvertent stimulus effects (for details see Wolsiefer, Westfall, & Judd, 2017). For this analysis we excluded participants with error rates across all blocks larger than 50% or who responded faster than 300 ms on at least 10% of all trials. We additionally discarded trials in which responses were faster than 400 ms or slower than 10 s. These exclusion criteria are the same as those used by Wolsiefer et al. (2017).

We analyzed standardized response latencies, that is, the time that elapsed between stimulus presentation and *correct* response divided by the standard deviation of all response latencies in a given block, Figure 1. To assess the reversal of the response mapping effect, we contrasted the common response mapping of Bob and negative words with the common mapping of Bob and positive words. Hence, larger values represent more favorable implicit evaluations.

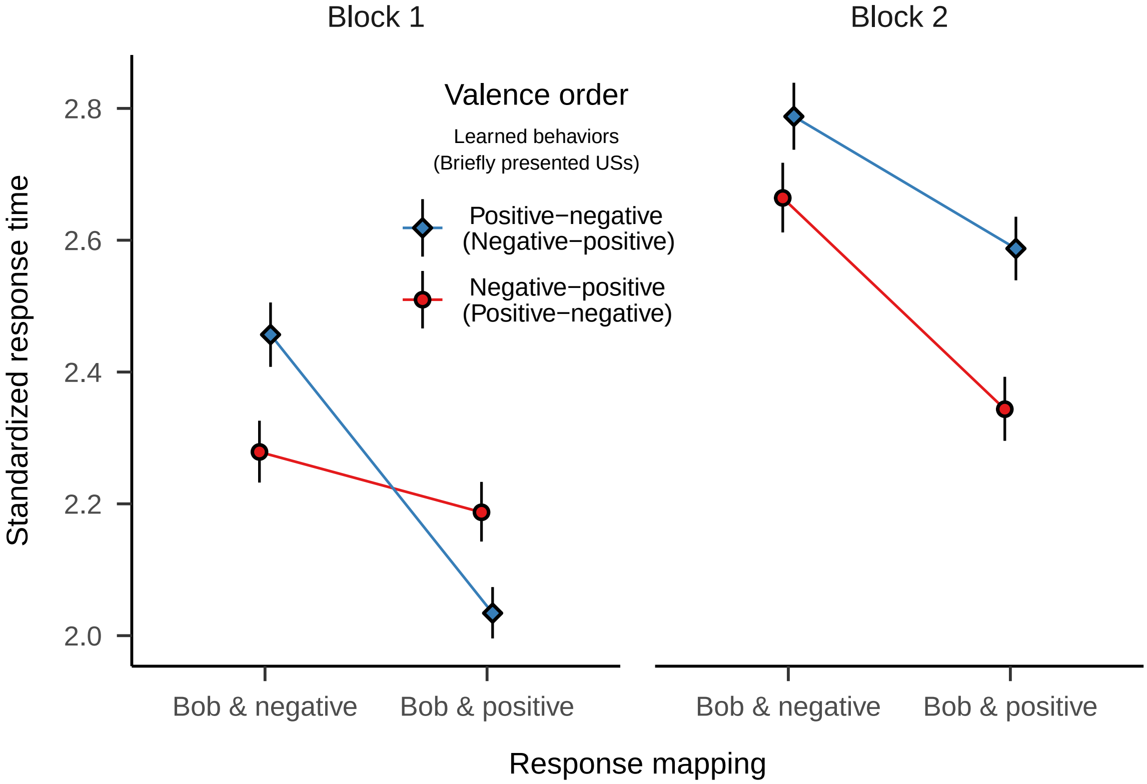

Figure 1. Standardized IAT response latencies across learning blocks. Black-rimmed points represent condition means, error bars represent 95% bootstrap confidence intervals based on 10,000 samples.

Table 2

Fixed effect estimates of the linear mixed model analysis of standardized IAT response times.

Effect	*b*	SE	*t*	*df*	*p*
Intercept	2.41	0.06	38.61	166.70	< .001
Response mapping	0.13	0.01	9.52	53.65	< .001
Learning block	-0.18	0.04	-4.95	151.66	< .001
Valence order	-0.04	0.06	-0.66	154.17	.510
Category	-0.17	0.02	-7.47	19.63	< .001
Word type	0.05	0.02	2.10	23.08	.047
Image type	-0.10	0.01	-11.02	19,355.13	< .001
Response mapping × Learning block	0.00	0.01	-0.43	58.55	.667
Response mapping × Valence order	-0.02	0.01	-1.69	62.83	.096
Learning block × Valence order	0.04	0.04	1.07	151.95	.288
Response mapping × Category	-0.01	0.01	-1.62	11.28	.133
Response mapping × Word type	0.00	0.01	0.22	28.39	.826
Response mapping × Image type	-0.04	0.01	-3.79	14,910.64	< .001
Learning block × Category	0.00	0.01	0.34	15.40	.741
Learning block × Word type	0.00	0.01	-0.51	47.64	.614
Learning block × Image type	0.01	0.01	1.12	12,077.70	.262
Valence order × Category	0.00	0.01	0.36	14.52	.724
Valence order × Word type	0.00	0.01	0.17	31.36	.865
Valence order × Image type	-0.01	0.01	-1.58	17,715.00	.114
Response mapping × Learning block × Valence order	-0.06	0.01	-7.02	82.01	< .001
Response mapping × Learning block × Category	0.00	0.01	-0.16	41.76	.875
Response mapping × Learning block × Word type	0.01	0.01	1.26	148.73	.209
Response mapping × Learning block × Image type	0.00	0.01	0.53	19,121.39	.599
Response mapping × Valence order × Category	-0.01	0.01	-0.62	15.41	.542
Response mapping × Valence order × Word type	0.01	0.01	0.73	35.54	.471
Response mapping × Valence order × Image type	0.00	0.01	-0.15	15,333.52	.878
Learning block × Valence order × Category	0.01	0.01	0.73	34.10	.469

Learning block × Valence order × Word type	0.00	0.01	-0.02	118.22	.986
Learning block × Valence order × Image type	-0.01	0.01	-1.01	12,888.28	.310
Response mapping × Learning block × Valence order × Category	0.02	0.01	2.37	76.83	.020
Response mapping × Learning block × Valence order × Word type	-0.01	0.01	-0.90	303.36	.371
Response mapping × Learning block × Valence order × Image type	0.00	0.01	0.13	18,425.77	.897

Note. The model additionally included random participant and item effects with random intercepts and random slopes for all manipulations during the learning procedure and their interactions.

Table 3

Random effect estimates and correlations of the linear mixed model analysis of standardized IAT response times.

	% of variance	1.	2.	3.	4.	5.	6.	7.	8.
Participant									
1. Intercept	.69	0.72	-0.04	-0.24	0.15				
2. Response mapping	.02		0.13	-0.09	0.21				
3. Learning block	.26			0.44	0.06				
4. Response mapping × Learning block	.00				0.06				
Stimulus									
1. Intercept	.01	0.09	-0.49	0.32	0.05	-0.78	0.66	0.60	0.12
2. Response mapping	.00		0.02	0.20	-0.39	0.79	-0.32	-0.89	-0.58
3. Learning block	.00			0.02	0.55	-0.35	0.68	-0.24	-0.76
4. Valence order	.00				0.03	-0.61	0.67	0.32	-0.10
5. Response mapping × Learning block	.00					0.01	-0.82	-0.82	-0.27
6. Response mapping × Valence order	.00						0.03	0.49	-0.03
7. Learning block × Valence order	.00							0.01	0.77
8. Response mapping × Learning block × Valence order	.00								0.01

Note. We report the estimated standard deviations in the main diagonals and the correlations in the off-diagnoals. The percentages of variance for the random effects were calculated by dividing each variance component by the total random variance, i.e., the sum of the random-effect variances.

Table 4

Post-hoc tests of changes in response mapping effects across blocks separately for pictures and words for standardized IAT response times.

Valence order	ΔM	95% CI	t	*df*	p
Pictures					
Negative-positive	-0.18	$[-0.32, -0.04]$	-2.96	43.31	.010
Positive-negative	0.14	$[-0.01, 0.29]$	2.26	32.00	.061
Words					
Negative-positive	-0.30	$[-0.43, -0.17]$	-5.22	198.96	$< .001$
Positive-negative	0.28	$[0.15, 0.41]$	4.81	160.94	$< .001$

Note. p values were Tukey-corrected for two comparisons.

In line with the ANOVA results, we found the expected three-way interaction between *Response mapping*, *Valence order*, and *Learning block*; the interaction was moderated by the type of stimulus that participants responded to (pictures of Bob and non-Bobs vs. positive and negative words; *Category*), Table 2 and 3. The three-way interaction prompted us to test the differences between response mapping effects in the first and second learning block for each valence order.

In line with the conventional ANOVA analysis, we found that response time differences suggested more favorable evaluations of Bob after the first than after the second block when the learned behaviors were first positive and later negative, $\Delta M = 0.21$, 95% CI [0.12, 0.30], $t(61.28) = 4.52$, $p < .001$. Vice versa, response time differences suggested more favorable evaluations after the second than after the first block when descriptions of Bob were first negative and later positive, $\Delta M = -0.24$, 95% CI $[-0.33, -0.15]$, $t(74.86) = -5.25$, $p < .001$. Again, these results indicate that the explicit evaluations and IAT scores were consistent.

Due to the significant four-way interaction, we additionally explored these contrasts separately for responses to pictures of Bob vs. non-Bobs and positive vs. negative words, Table 4. We found consistent changes in response mapping effects for both pictures and words, albeit the effects were larger for words.

Bayesian model comparison

We implemented the unconstrained model as a hierarchical linear model that encompasses each of the other models as special cases:

$$\begin{aligned}\hat{y}_{ijk} = &\mu + \nu_i + \eta_l x_{1il} + \\ &(\alpha + \tau_l x_{1il}) x_{2j} x_{3k} + \\ &(\beta + \upsilon_l x_{1il})(1 - x_{2j}) x_{3k}\end{aligned}$$

The model predicts the ith participant's response to evaluation measure j in the experimental block k. Responses are predicted as a combination of a grand mean μ, random participant intercepts ν_i (i.e., habitually higher or lower evaluations), a main effect of the labs η_l, and simple effects of learning block for rating scores (α) and IAT score (β). Additionally, we allowed the simple effects to be moderated by the labs (τ_l and υ_l represent the lab-specific deviations from the overall simple effects). The model does not include a main effect of evaluative measure because any mean differences between evaluative measures were leveled by the by-measure z standardization. x_{1il} represents l effect coded variables that indicate which lab participant i belongs to; x_{2j} indicates the evaluative measure (1 for rating score and 0 for IAT score), such that $\alpha + \tau_l$ is only relevant for rating scores and $\beta + \upsilon_l$ is only relevant for IAT scores; x_{3k} is an effect coded variable that is set to 0.5 for block 1 and -0.5 for block 2.

This model allowed us to place priors on the simple effects (in units of standardized mean differences d) for each evaluative measure and implement the theoretically motivated order constraints:

$$\begin{aligned}
\mathcal{M}_{\text{No effect}}: \;& \delta_\alpha = 0 \\
& \delta_\beta = 0 \\
\mathcal{M}_{\text{One mind}}: \;& \delta_\alpha \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2) \\
& \delta_\beta \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2) \\
\mathcal{M}_{\text{Two minds}}: \;& \delta_\alpha \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2) \\
& \delta_\beta \sim \text{Negative-Half-Cauchy}(r = \sqrt{2}/2) \\
\mathcal{M}_{\text{Any effect}}: \;& \delta_\alpha \sim \text{Cauchy}(r = \sqrt{2}/2) \\
& \delta_\beta \sim \text{Cauchy}(r = \sqrt{2}/2)
\end{aligned}$$

Additionally, we placed default multivariate Cauchy priors ($r = \sqrt{2}/2$) on lab main effects η_l as well as on lab effects on evaluative differences between blocks for rating scores (τ_l) and IAT scores (υ_l).

To formally assess whether the data from all labs exhibited consistent effects we added another model that enforced the order constraint of $\mathcal{M}_{\text{One mind}}$ and $\mathcal{M}_{\text{Two minds}}$ not only for the average block effects (α and β) but for each lab individually (i.e., $\alpha_l = \alpha + \tau_l$ and $\beta_l = \beta + \upsilon_l$; $\mathcal{M}_{\text{One mind everywhere}}$ and $\mathcal{M}_{\text{Two minds everywhere}}$).

For the analyses we drew 1 million samples to estimate the postrior distribtution of model parameters. Because the draws from the posterior distribution are used to estimate the Bayes factors for model comparisons that involve order constraints (Klugkist et al., 2005b), the number of draws implies upper and lower bounds on some of the reported Bayes factors. Most notably, as a direct consequence of the number MCMC samples the $\text{BF}_{\mathcal{M}_{\text{One mind}}/\mathcal{M}_{\text{Two minds}}} \in [\frac{1}{1\times10^6}, 1 \times 10^6]$.

Prior sensitivity analysis. Bayesian model comparison by Bayes factors are by definition sensitive to the specified prior distributions. To ensure that our inference is not contigent on our choice of piors we conducted prior sensitivity analyses for our key results.

Table 5

Results of the prior sensitivity analysis for the Bayesian model comparisons of primary interest.

r_α	r_β	$\text{BF}_{\mathcal{M}_{\text{One mind}}/\mathcal{M}_{\text{Two minds}}}$	$\text{BF}_{\mathcal{M}_{\text{One mind}}/\mathcal{M}_{\text{Any effect}}}$
0.50	0.35	1.00×10^6	4.00
0.96	0.35	1.00×10^6	4.00
0.96	0.53	1.00×10^6	4.00
0.96	0.71	1.00×10^6	4.00
1.41	0.35	1.00×10^6	4.00
1.41	0.53	1.00×10^6	4.00
1.41	0.71	1.00×10^6	4.00

Note. The Bayes factor (BF) in favor of $\mathcal{M}_{\text{One mind}}$ relative to $\mathcal{M}_{\text{Any effect}}$ is bounded within the range of $[0, 4]$ (see footnote 1 in the main article). r_α and r_β denote the scale for the Cauchy prior on the simple effects of learning block for rating scores (α) and IAT scores (β), respectively (in units of standard deviations).

Explicit and implicit measures.

Our choice of piors for the simple effects of learning block for rating scores (α) and IAT score (β) could be viewed as either overly optimistic or pessimistic. The prior on simple rating score effects places considerable probability mass on effects $d < 0.707$ although the previously reported effects were very large. Similarly, placing the same prior on the simple effects for rating and IAT scores could be criticized because the previously reported IAT score effects were considerably smaller than those of rating scores.

We, therefore, varied the scale for the Cauchy priors on the simple effects in the ranges of $0.50 < r_\alpha < 1.41$ and $0.35 < r_\beta < 0.71$ for rating and IAT scores, respectively. Considering results previous studies, we limited our reanalysis to combinations where the prior scale was larger for rating than for IAT effects. The results of the prior sensitvity analysis reassure us that our inference is robust to a wide range and combination of scales of the default Cauchy priors, see Table 5. The Bayes factors were not affected by the scale of the priors to any meaningful degree. This is because our data are informative enough to overwhelm the priors and because these Bayes factors primarily depend on the shape and location of the posterior distribution, not the prior distributions (Klugkist et al., 2005a).

Recognition task.

To test the robustness of our inference regarding participants recognition accuracy we varied the scale r of the Cauchy prior in a wide interval of $[0.50, 1]$. The resulting Bayes factors were $3.89 \times 10^6 < \text{BF}_{10} < 4.91 \times 10^6$ and thus varied by a factor of 1.26. These results again reassure that our inference is robust to a wide range of scales of the default Cauchy prior.

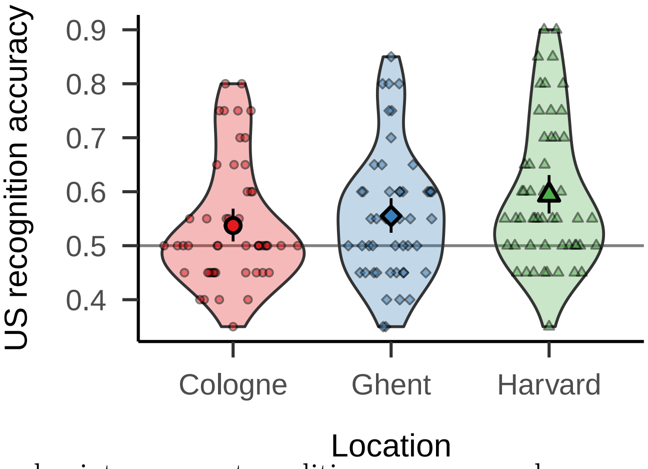

Figure 2. Black-rimmed points represent condition means, error bars represent 95% bootstrap confidence intervals based on 10,000 samples. Small points represent individual participants' accuracy. Violins represent kernel density estimates of sample distributions.

Prime recogniton and implicit evaluations

In contrast to the original results reported by Rydell, McConnell, Mackie, and Strain (2006), US recognition accuracy in this study was above chance, Figure 2. Memory for USs may, thus, have interfered with the associative learning process and prevented the predicted reversal of the IAT score differences. We, therefore, performed an exploratory regression analysis of US recognition and the IAT score difference between blocks used in the Bayesian analysis above. Positive values represent a more favorable evaluation after the block in which Bob was paired with positive learned behaviors and briefly presented negative USs. Conversely, negative IAT score difference between blocks indicate that the IAT effects reflect the valence of the briefly presented USs. If US recognition indeed obstructed the associative learning process, we would expect to observe a positive relationship between US recognition accuracy and IAT score differences between blocks: When US recognition is high, IAT score differences should reflect the valence of the learned behaviors but not with the US valence. We would expect to observe smaller and eventually negative IAT score differences as US recognition accuracy declines and associative learning takes over.

However, we were unable to detect any relationship between US recognition accuracy and IAT score differences between blocks, $b = -0.35$, 95% CI $[-1.63, 0.93]$, $t(151) = -0.54$, $p = .588$; the data even provide some evidence against such a relationship, $\mathrm{BF}_{01} = 5.01$. We centered US recognition at .5 and found that the intercept of the regression line was greater than zero, which indicates a positive IAT score difference despite at-chance US recognition accuracy, $b = 0.56$, 95% CI $[0.38, 0.74]$, $t(151) = 6.24$, $p < .001$. Hence, even for participants who exhibited no memory for briefly presented USs, IAT score differences reflected the valence of the learned behaviors, see Figure 3. These results provide no indication that the deviation of our findings from those reported by Rydell et al. (2006) are attributable to the above-chance US recognition accuracy in this study.

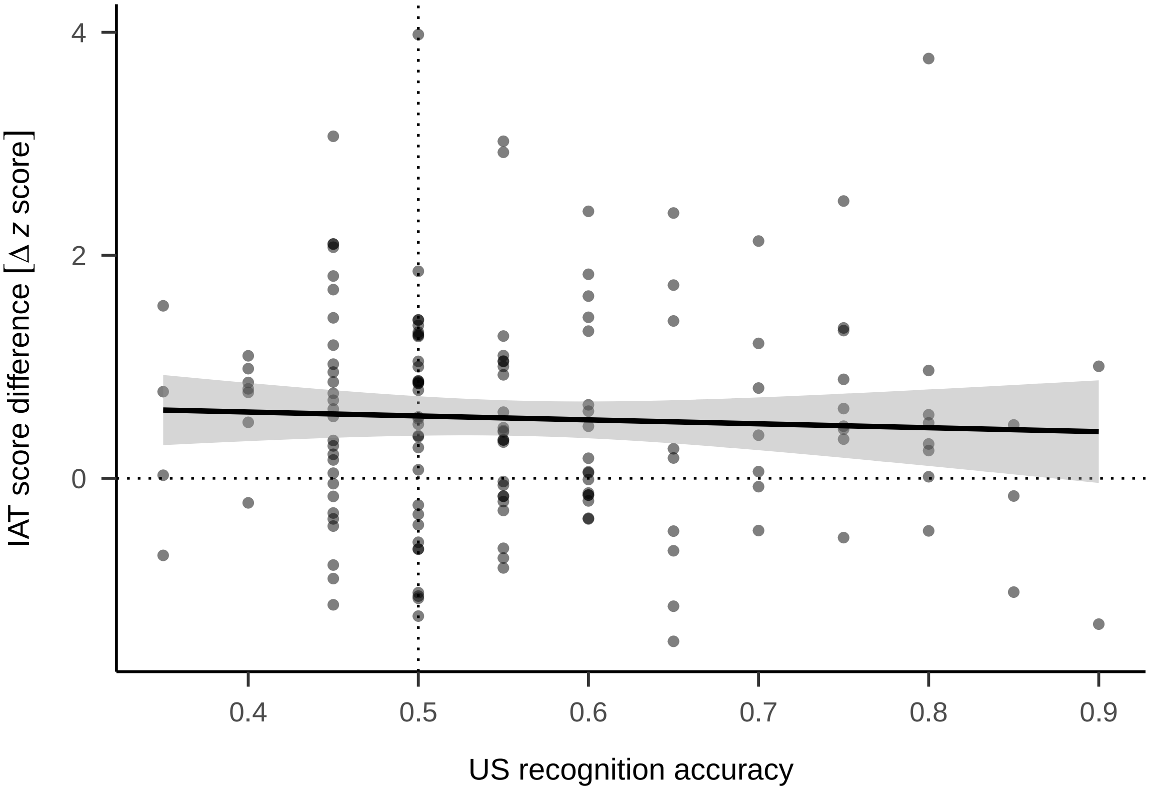

Figure 3. Scatterplot of prime recognition accuracy and evaluative differences in IAT scores between blocks in which Bob was presented with positive descriptions and those in which he was paired with negative descriptions.

References

Klugkist, I., Kato, B., & Hoijtink, H. (2005a). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*(1), 57–69. https://doi.org/10.1111/j.1467-9574.2005.00279.x

Klugkist, I., Laudy, O., & Hoijtink, H. (2005b). Inequality Constrained Analysis of Variance: A Bayesian Approach. *Psychological Methods*, *10*(4), 477–493. https://doi.org/10.1037/1082-989X.10.4.477

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of Two Minds: Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological Science*, *17*(11), 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x

Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods*, *49*(4), 1193–1209. https://doi.org/10.3758/s13428-016-0779-0