Supplementary Materials

Are taboo words simply more arousing?

Our study was not designed to compare potential differences on memory other than arousal between negative and taboo words. Thus, our data cannot explain why then (other than due to differences in arousal) negative words were different from taboo words in their locus of action on memory, i.e., association- versus item-memory. However, we aimed to clarify this question by conducting supplementary multidimensional scaling (MDS) analysis on word norms that were published by Janschewitz (2008) after we had carried out part of the present experiments. This database contains subjective ratings on seven dimensions (personal use, familiarity, imageability, arousal, valence, tabooness, and offensiveness) for 460 words pre-selected to be "taboo" (*n* = 92), "emotionally valenced," subdivided into four sets by high vs. low arousal and positive vs. negative valence (*n* = 46 each) or "emotionally neutral," subdivided into a related set, all within a single category (household objects and activities) and an unrelated set (*n* = 92 each). Euclidean distance matrices of proximities between the words were computed among all 460 words first and this input was used in PROXSCAL multidimensional scaling (MDS; implemented in SPSS), using the Kruskal loss estimation method (Kruskal, 1964). Stress, a badness-of-fit index, summarizes the adequacy of multidimensional solutions (Kruskal, 1964). A stress value of zero indicates a perfect fit of the dimensional configuration to the data, and a solution with a stress value of .10 is considered to adequately represent the data. The stress value for a one-dimensional solution was .38, for a two-dimensional solution it was .17, and for a 3-dimensional solution, the stress-value was .09, i.e., a value that can be considered reasonable for determining adequate dimensionality. The 3-dimensional solution was obtained in 25 iterations after which consecutive normalized raw stress values no longer substantially decreased (convergence criterion: normalized raw stress improvement > .0001). Variance explained by this solution was 99.23% (dispersion accounted for, D.A.F).

Pearson correlations between the original ratings and each word's common space coordinates (*Z*-transformed on word ratings) in the extracted dimensions are presented in Table S1. All seven original word ratings were highly correlated with Dimension 1, but the correlations were most pronounced for tabooness and offensiveness, followed by (negative) valence and arousal. Dimension 2 was also highly correlated with all ratings, but dominated by ratings of personal use and familiarity. Dimension 3 was dominated by imageability.

Table S1. Correlations between word ratings from Janschewitz (2008) and common space scores in three underlying dimensions derived from multidimensional scaling on the ratings (*n* = 460 words).

	Dimension 1	**Dimension 2**	**Dimension 3**
Personal Use	.59**	.77**	-.01
Familiarity	.32**	.92**	.05
Imageability	.51**	-.16**	.77**
Arousal	-.74**	.40**	.03
Valence[1]	.72**	-.13**	-.22**
Tabooness	-.91**	.18**	.17**
Offensiveness	-.90**	.21**	.12*

** $p < 0.01$ level, * $p < 0.05$ level; [1] High scores in "valence" represent positive valence.

To visualize how arousal or any other of the word ratings clustered the word types, we then conducted a series of regression analyses with the common space scores (only using Dimension 1 and Dimension 2) as predictors on each of the original ratings. These are outlined in Table S2. Regression vectors, based on the averaged beta weights from these analyses, were then plotted into the MDS space of Dimension 1 and Dimension 2 for each of the seven word attributes.

Table S2. Regression analyses predicting original word ratings from Janschewitz (2008) from Dimension 1 and Dimension 2 common space scores.

	R^2	F (2,457)	p	Beta	t (457)	p
Personal use	0.941	3643.650	<0.001			
Dimension 1				.594	52.279	<0.001
Dimension 2				.767	67.485	<0.001
Familiarity	0.947	4290.185	<0.001			
Dimension 1				.324	30.778	<0.001
Dimension 2				.919	87.368	<0.001
Imageability	0.285	91.048	<0.001			
Dimension 1				.509	12.856	<0.001
Dimension 2				-.162	-4.102	<0.001
Arousal	0.71	547.044	<0.001			
Dimension 1				-.741	-29.182	<0.001
Dimension 2				.395	15.572	<0.001
Valence	0.532	259.879	<0.001			
Dimension 1				.717	22.422	<0.001
Dimension 2				-.132	-4.122	<0.001
Tabooness	0.854	1336.491	<0.001			
Dimension 1				-.906	-50.691	<0.001
Dimension 2				.182	10.169	<0.001
Offensiveness	0.845	1244.595	<0.001			
Dimension 1				-.895	-48.562	<0.001
Dimension 2				.211	11.444	<0.001

As can be seen in Figure S1, Dimension 1 differentiated well between taboo words and other types of words, while not differentiating well between high and low arousing negative words as well as between positive words and both types of neutral words. Importantly, if arousal alone had separated taboo words from the other word types, one would have expected a different pattern: Neutral, low arousing negative, and low arousing positive words should have ranked highest, followed by high arousing negative and high arousing positive words, and taboo words ranking lowest.

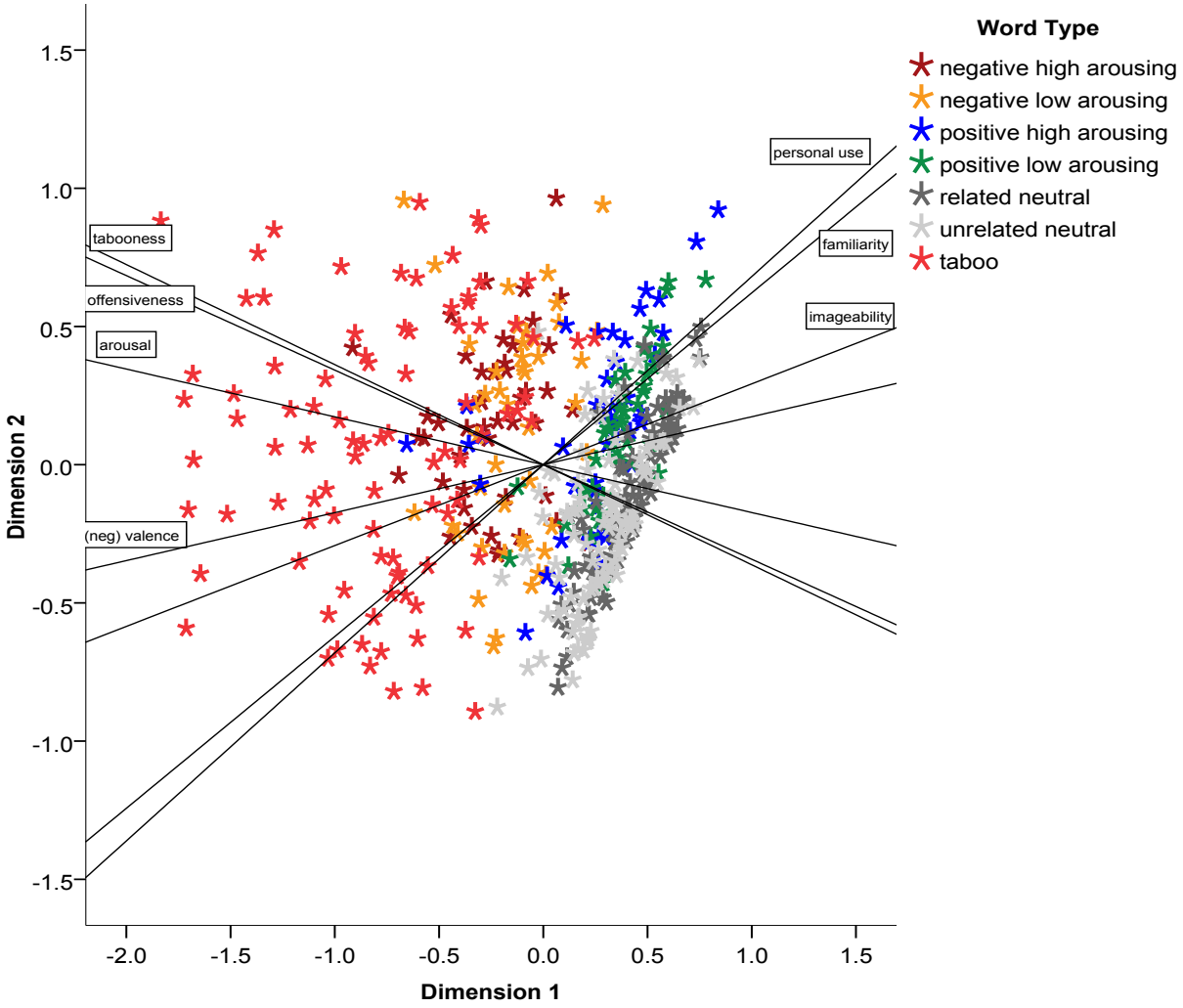

Figure S1: Placements of the 460 words in MDS Dimension 1 and Dimension 2, separated by word type. Averaged beta scores were used to plot regression vectors onto the original word ratings.

As a different way to characterize which word types differed from each other in their localization on the MDS dimensions, we then conducted three ANOVAs on the common space scores in each of the three dimensions with word type as a factor. For Dimension 1, Levene's test for homogeneity of variances indicated inhomogeneity between word types [$W(6,453) = 20.41$, $p < 0.001$]. Adjusting the degrees of freedom, the Welch test indicated

significant differences between the word types [$F(6,171) = 144.3$, $p < 0.001$]. As illustrated in Figure S2, post-hoc *t*-tests (corrected for multiple comparisons using the Games-Howell test, assuming unequal variances) showed significant differences between three clusters of word types: Taboo words ranked significantly lower in Dimension 1 than the two negative word types [$t(129) = 11.1$, $p < 0.001$, corrected] as well as positive and neutral words [$t(106) = 21.9$, $p < 0.001$]. Negative words also ranked lower than positive and neutral words [$t(154) = 19.02$, $p < 0.001$]. There were no significant differences between negative high-arousing and low-arousing words and no differences among positive high-arousing, positive low-arousing, and both types of neutral words [all p's > .1, corrected].

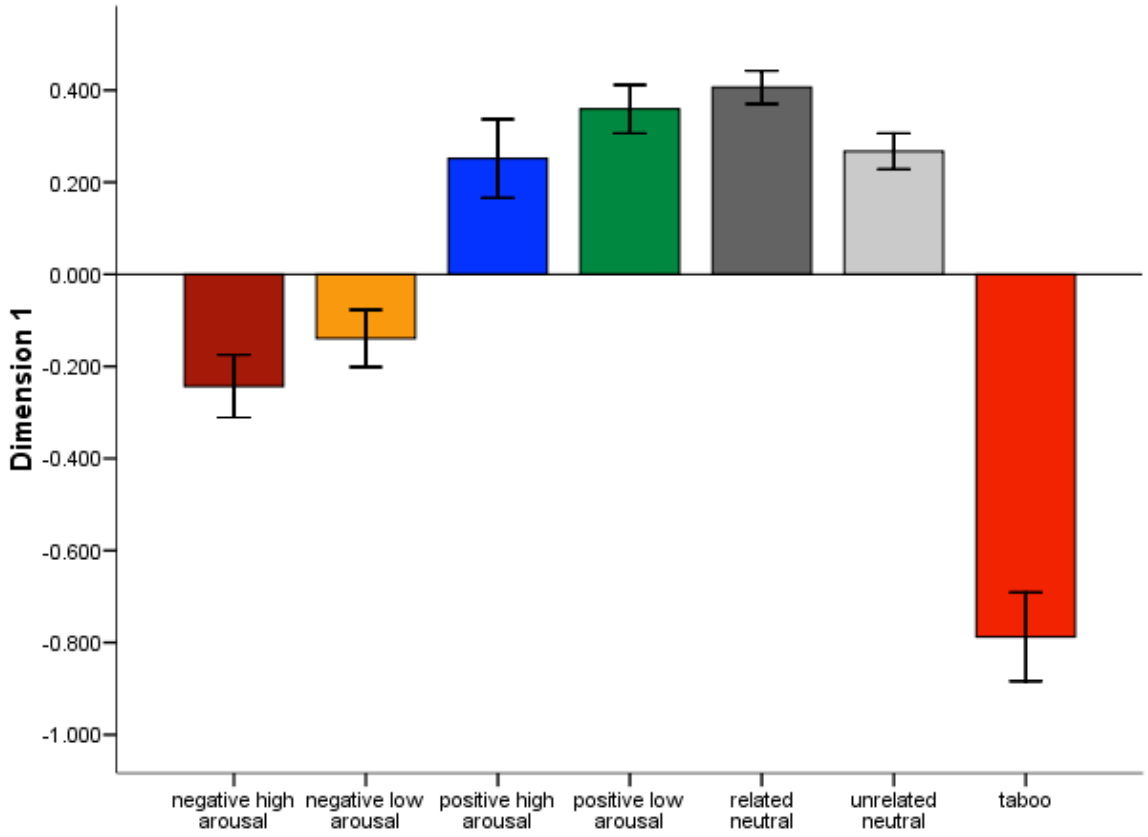

Figure S2: Average common space scores of words in Dimension 1, separated by pre-determined word types. Error bars represent 95% confidence intervals.

Thus, taboo words differed significantly from both types of negative words as well as from positive/neutral words along Dimension 1. This reiterates our visual interpretation of the scatter plot in S1: If arousal alone had separated taboo words from the other word types, one would have expected low arousing neutral, negative, and positive words scoring highest, followed by high arousing negative and positive words, and taboo words scoring lowest.

Analogous ANOVAs were conducted on common space scores between the pre-determined word types in Dimensions 2 and 3. For Dimension 2, Levene's test for homogeneity of variances indicated inhomogeneity between word types [$W(6,453) = 8.66$, $p < 0.001$]. Adjusting the degrees of freedom, the Welch test indicated significant differences between the word types [$F(6,172) = 15.16$, $p < 0.001$]. As can be seen in Figure S3, the two types of neutral

words were significantly different from all other word types in the post-hocs [$t(405) = 9.11$, $p < 0.01$, corrected with Games-Howell test]. The two neutral word types did not differ significantly from each other [$t(182) = 2.34$, $p = 0.23$]. Importantly, taboo words were not significantly different from any word type except the two types of neutral words [$t(128) = 4.28$, $p < 0.05$].

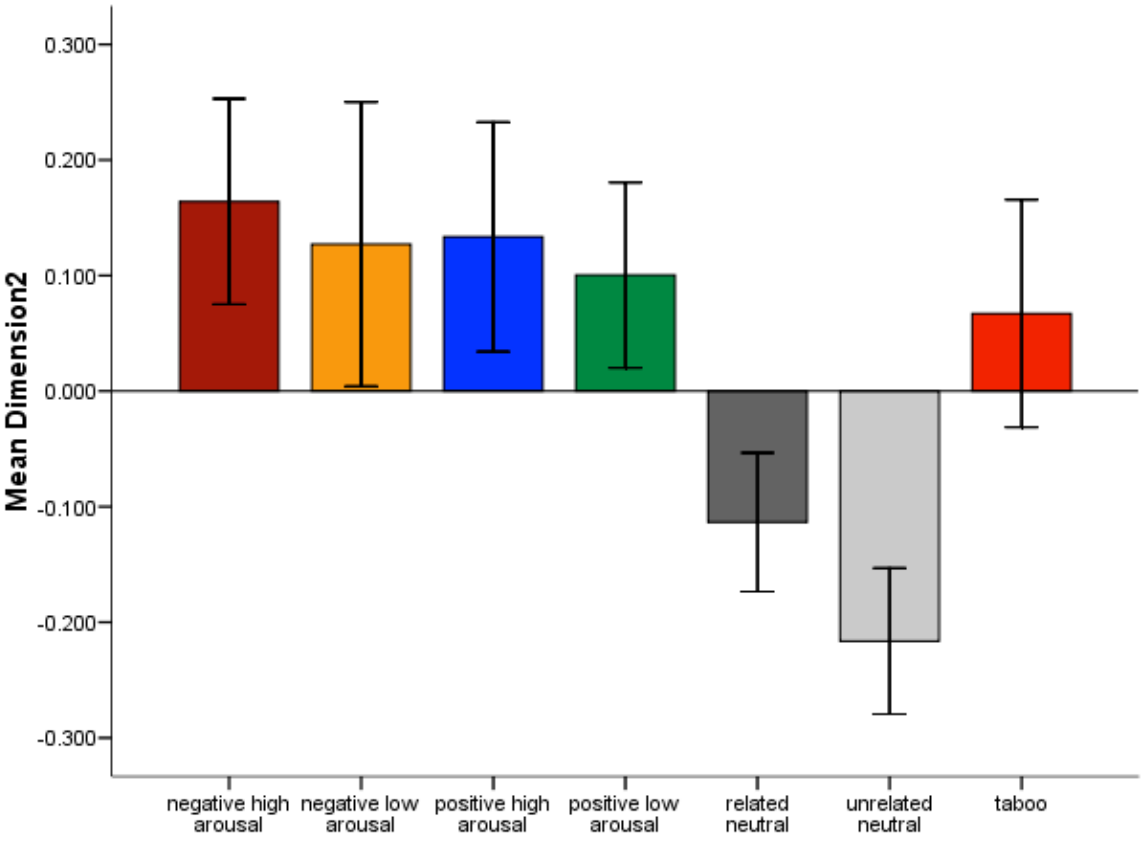

Figure S3. Average common space scores of words in Dimension 2, separated by pre-determined word types. Error bars represent 95% confidence intervals.

For Dimension 3, word types also differed significantly from each other [$F(6,453) = 11.23$, $p < 0.001$]. Illustrated in Figure S4, positive words together with negative low arousing words were significantly different from all other word types [$t(453] = 6.78$, $p < 0.01$, corrected]. Importantly, taboo words were not significantly different from any word type except high arousing positive words [$t(453) = 5.77$, $p < 0.01$] and low arousing positive words [$t(453) = 4.82$, $p < 0.01$].

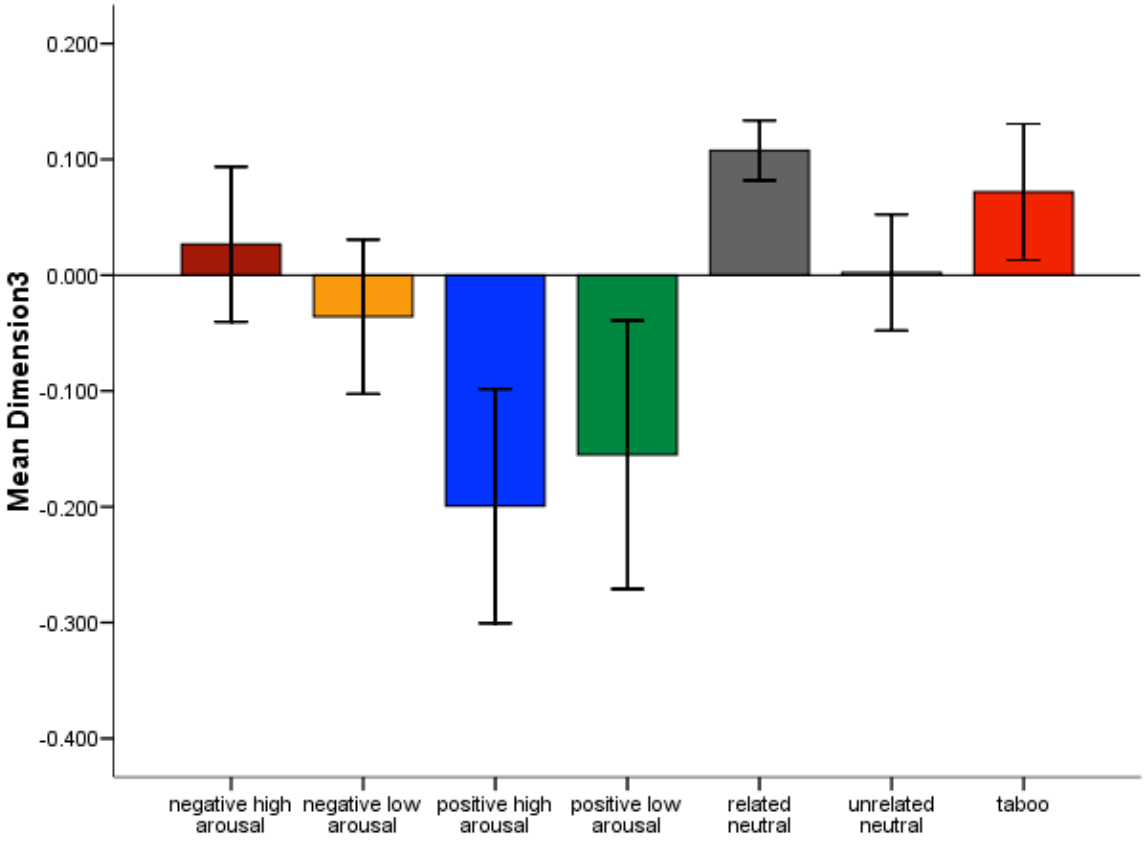

Figure S4. Average common space scores of words in Dimension 3, separated by pre-determined word types. Error bars represent 95% confidence intervals.

References

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behaviour Research Methods, 40*, 1065–1074.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika, 29*, 1–27.