Journal of Memory and Language 66 (2012) 695–716

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Emotional arousal does not enhance association-memory

Christopher R. Madan [a,*], Jeremy B. Caplan [a,b], Christine S.M. Lau [a,c], Esther Fujiwara [b,c]

[a] Department of Psychology, University of Alberta, Edmonton, AB, Canada
[b] Department of Centre for Neuroscience, University of Alberta, Edmonton, AB, Canada
[c] Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

ARTICLE INFO

Article history:
Received 4 July 2011
revision received 30 March 2012
Available online 28 April 2012

Keywords:
Arousal
Emotion
Association-memory
Taboo
Paired-associate learning
Cued recall

ABSTRACT

Emotionally arousing information is remembered better than neutral information. This enhancement effect has been shown for memory for items. In contrast, studies of association-memory have found both impairments and enhancements of association-memory by arousal. We aimed to resolve these conflicting results by using a cued-recall paradigm combined with a model-based data analysis method (Madan, Glaholt, & Caplan, 2010) that simultaneously obtains separate estimates of arousal effects on memory for associations and memory for items. Participants studied sequentially presented words in pairs that were pure (NEGATIVE–NEGATIVE or NEUTRAL–NEUTRAL) or mixed (NEGATIVE–NEUTRAL or NEUTRAL–NEGATIVE). Cued recall tests had NEUTRAL or NEGATIVE probes and NEUTRAL or NEGATIVE targets. We found impaired memory for associations involving negative words despite enhanced item-memory (more retrievable targets). A category-list control condition explained away the item-memory enhancement but could not explain the impairment of association-memory due to arousal. A second experiment with identical structure but using higher-arousing taboo words revealed increased cued recall of taboo than neutral words. However, this was exclusively mediated by item-memory effects with neither enhancement nor impairment of association-memory. Thus, cued recall was lower for pure negative pairs and higher for pure taboo pairs, but our modeling approach determined a different locus of action for these memory impairing or increasing effects: Although item memory was increased by arousal, association-memory was impaired by negative words and unaffected by taboo words. Our results suggest that previous results reporting an enhancement of association-memory due to arousal may have instead been solely driven by enhanced item-memory.

1. Introduction

Cued recall is often used as a test of association-memory. However, cued recall requires one not only to remember the relationship between items (association-memory) but also the items themselves (item-memory). Thus, verbal cued recall is influenced by item-memory for the probe and target words in addition to (or even instead of) memory for the association itself. Single-word properties can act at any of these levels. We recently demonstrated this with the word properties of imageability and word frequency (Madan, Glaholt, & Caplan, 2010). Briefly, imageability is a measure of how conducive a word is to mental imagery (e.g., high-imageability: BIKE; low-imageability: CLAIM), whereas word frequency is a measure of the probability of occurrences of a word (e.g., high-frequency: AREA; low-frequency: MICA). We found that in both manipulations, cued recall accuracy was better for pure high pairs (pairs in which both words were high-imageability or high-frequency, respectively) than for pure low pairs (pairs in which both words were low-imageability or low-frequency,

* Corresponding author. Address: Department of Psychology, Biological Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. Fax: +1 780 492 0023.
E-mail address: cmadan@ualberta.ca (C.R. Madan).

0749-596X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2012.04.001

respectively). However, in "mixed" pairs, which consisted of one high item and one low item, we found different effects for each word property. In the imageability manipulation, accuracy for the mixed pairs was symmetrical regardless whether high/low words were probes or targets and fell between accuracy for the pure high and pure low pairs. In the word frequency manipulation, accuracy for the mixed pairs was asymmetric, with better performance when the cued recall target was a high-frequency word than if it was a low-frequency word. However, even by including all possible types of pairs, cued recall accuracy is still dependent on both item- and association-memory processes. To systematically disentangle the influences of item- and association-memory producing these patterns of results across manipulations, we then applied a model-based approach. This approach simultaneously obtains formal estimates of the influence of single-item properties (i.e., imageability and word frequency) on item- and association-memory. By including all possible probe and target combinations (i.e., high–high, high–low, low–high, low–low), we were able to obtain estimates of the cued recall target's retrievability (model parameter: t), the probe word's effectiveness (model parameter: p), and strength of the memory for the relationship between the two items (model parameters: r_1, reflecting any difference in memory between high–high and mixed pairs, and r_2, reflecting any difference in memory between mixed and low–low pairs.) Both relationship model parameters depend only on pair type but not upon which item was used as the probe or target in cued recall. Thus, one could imagine that memorizing pairs of dissimilar items (i.e., mixed pairs) is more challenging than memorizing pairs of similar items. Our model can test then, if and how the item manipulation (e.g., high/low imageability) interacts with such basic differences between pure and mixed pair recall. For each of these parameters, a value greater than one signifies that the process is enhanced due to the manipulation, whereas a value below one signifies an impairment of that process due to the manipulation. (For more detail on our modeling approach, see page 17.) Despite finding no difference between word frequency and imageability in cued recall performance in pure pairs, with our modeling approach we found that imageability primarily enhanced association-memory (r_1 and $r_2 > 1$) and that word frequency primarily enhanced target retrievability ($t > 1$; also reported by Criss, Aue, & Smith, 2011). We therefore extend the use of this cued recall paradigm and modeling approach to an important topic of memory research: the influence of arousal on association-memory.

Previous research has shown that emotional arousal enhances memory for single items or events, a finding that has been replicated with many different paradigms and materials, including arousing words and pictures, as well as more realistic events like flashbulb memories (e.g., Berntsen & Rubin, 2004; Bradley, Greenwald, Petry, & Lang, 1992; Brown & Kulik, 1977; Christianson, 1992; Mather & Sutherland, 2011). Previous studies of arousal effects on association-memory have used a variety of paradigms in which two elements (typically an arousing item paired with a neutral item) had to be bound together in memory. If those elements coincide, association-memory is often enhanced (e.g., font color of a presented word; Doerksen & Shimamura, 2001; Kensinger & Corkin, 2003, but see Davidson, McFarland, & Glisky, 2006). If the two elements do not coincide, association-memory appears to be reduced due to arousal (e.g., peripheral neutral line drawings in the presence of central arousing scene pictures; Touryan, Marian, & Shimamura, 2007). These findings have been derived from paradigms with a number of potentially problematic characteristics. Almost all previous studies in this area used incidental encoding instructions and dissimilar items to-be-bound in memory. Under these circumstances, several ill-controlled factors could account for a differential formation of associations with arousing compared to non-arousing items in memory. For example, when presenting an arousing and neutral item together, attention will likely be drawn to the arousing item if there is no explicit instruction to encode the association between the two. This may then influence later memory for the neutral item. Such impairment may not reflect impaired binding of elements in memory *per se*, but rather an effect of unequal attention to the two components, which may then impair encoding of the relation between them. An unequal attention distribution could be further amplified by using dissimilar assemblies of central arousing and peripheral neutral elements. Four published studies avoided some of these ambiguities by using intentional encoding instructions with pairs of items that were of the same kind (word–word pairs). To test memory for associations, two of them used associative recognition at retrieval (Onoda, Okamoto, & Yamawaki, 2009; Pierce & Kensinger, 2011), and two used cued recall (Guillet & Arndt, 2009; Zimmerman & Kelley, 2010). In Onoda et al. (2009), participants learned pairs of words that were either both negative or both neutral. Subsequently, on associative recognition tests, words were either presented with their original associate ("intact"), or with a word from another pair ("rearranged"). Participants were more accurate on neutral than negative pairs. Pierce and Kensinger (2011) presented participants with pairs of words that consisted of either two negative words, two neutral words, or two positive words. When tested immediately, memory for negative pairs was worse than for neutral pairs, as well as for positive pairs (Pierce & Kensinger, 2011 Experiment 1). When a longer study–test delay was used, memory was enhanced for negative compared to neutral/positive pairs (see Pierce & Kensinger, 2011 Experiment 2).

Arguably, associative recognition is a more direct test of association-memory than cued recall, and that it is unaffected by item-specific information. However, recent evidence has shown that associative recognition is in fact also susceptible to manipulations of item strength (Buchler, Light, & Reder, 2008; Criss & Shiffrin, 2005; Kelley & Wixted, 2001). Thus, it is possible that prior findings of an impairment of association-memory due to arousal may instead have been primarily driven by arousal effects on item-memory, even in associative recognition. In this case we will expect that more arousing pairs will have higher false alarm rates than neutral pairs, rather than a reduced hit rate. Indeed, Pierce and Kensinger (2011) found that negative pairs had higher false alarm rates than neutral pairs. False alarm rates were also higher for positive pairs than for neutral pairs, but this difference was not statistically significant. Nonetheless, hit rates were equivalent across all three pair types (unfortunately, Onoda et al., 2009, did not report hit rates and false alarm rates in their associative recognition results, but only an accuracy measure that collapsed the two rates). Taken together, neither associative recognition nor cued recall is a direct test of association-memory; they can both be influenced by item-memory effects. By applying the same experimental design we have used in Madan et al. (2010), here using manipulations of arousal, we are able to systematically disentangle the effects of arousal on item- and association-memory on cued recall performance.

Using cued recall at test instead of associative recognition, Guillet and Arndt (2009) conducted a paired-associate study where words were presented in central and peripheral locations during study (Experiments 2A–2C). The central word was always used as the cued recall probe, and was either a neutral, negative, or taboo word. The peripheral word was always later used as the cued recall target, and was always neutral. Guillet and Arndt (2009) found better cued recall performance with taboo probe words, as compared to neutral or negative probe words. This was interpreted as an enhancement of association-memory due to arousal. In contrast, Zimmerman and Kelley (2010) had participants study pairs of words that consisted of either two negative words or two neutral words. Compared to neutral pairs, they found enhanced free recall (testing item-memory) for single negative words, but impaired cued recall (aimed at testing association-memory) for negative pairs. Participants also made judgements of learning (JOLs) during study, and significantly overestimated their ability to remember the negative pairs relative to overt cued recall performance, while their JOLs for neutral pairs were more precise. Thus, the findings of Zimmerman and Kelley (2010) suggest that associations between two negative items are impaired, and that this may be due to a subjective overconfidence in memory when subjects encounter such pairs. This may possibly have caused subjects in Zimmerman and Kelley (2010) to apply less effort when learning negative pairs than neutral pairs.

We propose that these two cued recall studies (Guillet & Arndt, 2009; Zimmerman & Kelley, 2010) may have come to opposite conclusions because they did not use all possible pair types of arousing/neutral items, and the cued recall measures may have confounded effects of item- and association memory. To illustrate, Guillet and Arndt (2009) found enhanced cued recall of a neutral word that had been paired with a taboo word, compared to having been paired neutral word. It is possible that the results of Guillet and Arndt (2009) could also be produced if taboo words are simply better probe items (e.g., preferentially processed) than neutral words, without truly increasing memory for the association *per se*. In contrast, Zimmerman and Kelley (2010) reported impaired cued recall when both words were negative (versus both words being neutral). This reduced recall of negative pairs could exclusively result from non-association-memory mechanisms. Single arousing words could be easier to recall as targets in cued recall compared to neutral words, but they could be worse probes (for example, an arousing probe word could distract the participant momentarily and make it harder to search for the specific corresponding association in question, similar to working-memory impairments due to arousal; Mather et al., 2006). The net result of two such opposing effects could produce a reduction in cued recall (as Zimmerman & Kelley, 2010 observed), but could still be unrelated to association-memory *per se*. Importantly, by only using a subset of all possible combinations of pair arrangements, it is impossible to know if differences in memory effects observed in cued recall are driven by item- or association-memory effects. Furthermore, it can be especially difficult to compare across cued recall studies if they use a different subset of the possible pair types: Guillet and Arndt (2009) only manipulated the properties of the probe item, while always using a neutral target item. Zimmerman and Kelley (2010) instead manipulated the pair as a whole, where pairs were composed of either two arousing items or two neutral items, but never one of each.

In Experiment 1, we used a 2×2 design with all possible combinations of moderately arousing negative (N) and neutral (n) words as cued recall probes and targets (i.e., NN, Nn, nN, nn). We expected that by being able to better account for item-memory effects (model parameters for probe effectiveness, p, and target retrievability, t) our model-based analysis would provide more conclusive evidence regarding the influence of arousal on association-memory than has been previously demonstrated. Thus, we had three plausible alternative hypotheses: (a) arousal may enhance association-memory (as estimated by the model parameters for the strength of the relationship between the items, r_1 and r_2), (b) arousal may impair association-memory, or that (c) previously reported results regarding the effects of arousal on association-memory may instead be driven solely by item-memory effects (estimated by p and t).

We also included a control group that studied words composed of categorized neutral (C) and random neutral (n) words to test whether effects of negative arousing words could be explained in terms of category effects. Finally, in Experiment 2, we examined highly arousing word (taboo words [T]) to further test the influence of increased arousal on association-memory.

Additionally, all four of the studies discussed earlier used simultaneous presentation of both words in the pair, which may have resulted in ill-controlled differential attention to each one of the two words during encoding, which in turn may have caused differential recall rather than arousal of the word(s) *per se*. To avoid this potential confound, we instead presented the words of each pair sequentially to allow participants to equally and fully attend both.

Based on the results of these previous studies, we outline our three alternative hypotheses regarding the influence of emotional arousal on cued recall as follows:

1.1. Association-memory enhancement hypothesis

With an intentional encoding instruction and sequential presentation of the two words in each pair, participants' ability to fully attend to both words in each pair equally and to deliberately form memory for the association should be maximal. If arousal increases the formation of associative memories, as proposed by Guillet and Arndt (2009), that memory for the association should be greater when the pair involves one or more arousing words than when a pair involves only neutral words: both r_1 and $r_2 > 1$.

1.2. Association-memory impairment hypothesis

Alternatively, if participants apply less effort when forming associations that involve arousing words (as suggested by the results of Zimmerman & Kelley, 2010) distracting or otherwise damaging effects of arousal on the formation of association-memory cannot be eliminated by sequential presentation alone. Therefore, our alternative prediction is that memory for the association will be reduced when a pair involves one or more arousing words than when a pair involved only neutral words: both r_1 and $r_2 < 1$.

1.3. Item-memory hypothesis

Emotionally arousing items are remembered better than neutral items. If arousing items are more easily retrieved than neutral items, the results we summarized earlier that were based on subsets of possible pair types may instead have been driven solely by the modulation of probe effectiveness and/or target retrievability due to arousal. The prediction is that memory for associations will be unaffected by arousal (both r_1 and $r_2 = 1$) in our model), but instead may be driven by an impairment of probe effectiveness due to arousal ($p < 1$) along with an enhancement of target retrievability ($t > 1$).

2. Experiment 1

We tested cued recall of word pairs comprised of moderately arousing, negative words (N) and neutral (n) words. Previous research has suggested that item-memory for negative words may be enhanced because negative words have a closer semantic cohesiveness than 'random neutral' words (Buchanan, Etzel, Adolphs, & Tranel, 2006; Talmi & Moscovitch, 2004). If the enhancement of item-memory due to arousal can be largely explained by semantic cohesiveness, the same may hold for association-memory. To rule out this possibility, we also investigated memory performance with categorized neutral (C) and random neutral n) words in a separate group of participants.

To identify any asymmetries due to sequential presentation, we probed pairs both in the usual, "forward" direction (given the first item, recall the second item) and in the "backward" direction (given the second item, recall the first item). To illustrate with an example, a negative (N) word would be the target in neutral–negative (nN) pairs tested in the forward direction as well as in negative–neutral (Nn) pairs tested in the backward direction; however, the arousing word in each pair would be encoded in a different order due to the sequential presentation. Order effects may play a role in the influence of arousal on memory. For example, Miu, Heilman, Opre, and Miclea (2005) found stronger memory impairment for words immediately preceding, rather than following, emotional oddballs ('emotion-induced retrograde amnesia'; see also Strange, Hurlemann, & Dolan, 2003). Thus, by including both test directions we are also sensitive to possible temporal asymmetries during encoding.

3. Methods

Prior to Experiment 1, we conducted a norming study. This preliminary study was necessary as no published database contained sufficient normative data at the time of commencement of our experiments. We additionally used this norming study to obtain similarity judgements, which were essential for matching our negative and categorized neutral lists.

3.1. Norming study for Experiment 1

3.1.1. Participants

A total of twelve participants participated in our norming study. Five participants who took part in the study received partial credit towards an undergraduate psychology course and seven students received monetary payment of $12 for their participation. None of the participants from this norming study took part in the memory task of Experiments 1 or 2.

3.1.2. Methods

Participants rated a larger pool of words for arousal and valence using two Self-Assessment Manikins (SAM; Bradley & Lang, 1994) as well as the similarity between selected pairs of words. Specifically, participants were given pencil-and-paper questionnaires with 160 words to be judged on arousal and valence as well as 240 word pairs to be judged on semantic similarity. To avoid fatigue, tasks alternated between blocks of 40 arousal/valence judgments and 80 semantic similarity ratings.

Twelve word lists were created: (a) Four lists of 24 moderately arousing, negative words were preselected based on norms from the Affective Norms for English Words (ANEW; Bradley & Lang, 1999). According to this reference, each word has an assigned valence value (scored on a scale of 1 [very negative] to 9 [very positive]), and an arousal value (scored on a scale of 1 [not arousing] to 9 [highly arousing]). (b) For each of four categories 'driving', 'school', 'house', and 'business,' one list of categorized neutral words was created, resulting in four categorized lists. The categories 'driving' and 'school' were taken from Talmi and Moscovitch (2004) and 'house' and 'business' were self-generated, each containing 48 words. (c) Four lists of 48 random neutral words each were created.

For the arousal/valence ratings, lists were counterbalanced such that each participant rated a quarter of the word pool. Thus, across four participants, each word was rated once for arousal and valence. As our norming study had twelve participants, each word was rated by exactly three participants.

Due to the large number of possible semantic similarity judgments, each participant was randomly assigned 240 similarity judgments: 12 (negative) or 24 (categorized neutral or random neutral) word pairs from each of the twelve lists. This (incomplete) set of similarity ratings was used to quantify the similarity between a given word and the rest of the list.

Inter-rater reliability for was reasonably high for both arousal [Cronbach's $\alpha = .78$] and valence [$\alpha = .86$] ratings. Inter-rater reliability could not be calculated for semantic similarity judgements, since each participant was given different word pairings.

3.1.3. Word-selection criteria

To equate arousal levels among negative lists and ensure higher arousal for negative lists than for the neutral lists, we removed words from our negative lists based on our obtained arousal ratings. Likewise, to ensure that words within all categorized neutral and negative lists were perceived as equally related, we removed words that were less related to the rest of the words in the list. This ensured a narrow range of similarity ratings within each list. We further removed words from each list to ensure that the lists were equivalent for several other word properties known to affect memory performance: familiarity, imageability, word frequency, and number of letters (MRC Psycholinguistic Database; Wilson, 1988).

3.1.4. Final word pools

We produced three word pools for use in Experiment 1: 64 negative words, 64 categorized neutral words, and 64 random neutral words. See Table 1 for word property statistics for the specific words used in Experiment 1 (also see Appendix A).[1]

With the creation of these three word pools, we proceeded with our memory study.

3.2. Participants

A total of 240 undergraduate students (mean age ± *sd* = 19.02 ± 1.57; 88 males and 152 females) participated in our study for partial credit in an introductory psychology course at the University of Alberta. Data from seven of these participants were lost due to machine error. Participants were randomly assigned to either the Negative group ($N = 120$) or the Category group ($N = 113$). All participants were required to have learned English before the age of six and to be comfortable typing. Participants gave written informed consent prior to beginning the study, which was approved by a University of Alberta Research Ethics Board.

Table 1
Arousal, valence, and similarity ratings are subjective ratings from the norming task in Experiment 1. Familiarity, imageability, and frequency ratings were obtained from the MRC Psycholingustic Database (Wilson, 1988). Mean ratings are shown with standard deviation in parentheses. Means in a row with the same superscript are not significantly different at $p < .05$.

	Negative	Category	Neutral
Arousal	5.20 (0.91)^a	3.68 (0.82)^b	3.75 (0.58)^b
Valence	2.23 (1.00)^a	4.20 (0.69)^b	4.22 (0.54)^b
Similarity	4.78 (0.63)^a	4.75 (0.61)^a	2.92 (0.58)^b
Familiarity	520 (48)^c	531 (44)^c	528 (46)^c
Imageability	490 (69)^c	492 (79)^c	505 (104)^c
Frequency	44.5 (50.5)^c	58.2 (73.7)^c	54.6 (52.1)^c
Length	5.92 (1.60)^c	6.36 (1.80)^c	5.80 (1.56)^c

3.3. Materials

We used the three word pools produced in the norming study. Stimuli in the Negative group were negative and random neutral words; stimuli in the Category group were categorized and random neutral words.

3.4. Procedure

Each participant had one practice study set (which was not included in the data analysis) followed by eight experimental study sets. Each study set consisted of three phases: study phase, distractor phase, and cued recall. The session concluded with a final free recall task.

All stimuli were presented in a white "Courier New" font, which ensured fixed letter width, on a black background, in the center of the screen. Words were presented sequentially, for 3000 ms each, plus a 50 ms inter-stimulus interval within pairs and a 4000 ms inter-pair interval during which a fixation cross, "+", was displayed in the center of the screen. During the study phase, participants were presented with eight word pairs, asked to study the pairs, and told that their memory for the pairs would be tested later on. Each study set consisted of two pairs of each of the four pair types: NEGATIVE–NEGATIVE (NN), NEGATIVE–NEUTRAL (Nn), NEUTRAL–NEGATIVE (nN), and NEUTRAL–NEUTRAL (nn). Pair types in the Category group were: CATEGORIZED–CATEGORIZED (CC), CATEGORIZED–NEUTRAL (Cn), NEUTRAL–CATEGORIZED (nC), and NEUTRAL–NEUTRAL (nn). NN/CC/nn are 'pure' pairs and Nn/nN/Cn/nC are 'mixed' pairs. Word pairings, word membership by pair type, order of pairs, and order of pair types were all randomized across participants. For the Category group, within each study set (eight word pairs), all categorized neutral words belonged to the same category (e.g., all 'school' words). The category used in each study set of eight word pairs was randomized such that each of the first four study sets used a different semantic category.

The distractor task consisted of four simple arithmetic problems, in the form of A + B + C = _, where A, B, and C were randomly selected digits between two and eight.

[1] We also calculated LSA $cos(\theta)$ as an additional measure of within-pool word similarity (Landauer & Dumais, 1997). LSA $cos(\theta)$ within each of our word pools (mean ± *sd*) are as follows: negative (.230 ± .128), categorized (.153 ± .145), and neutral (.077 ± .088). Independent-sample *t*-tests (with *df* based on the effective number of independent comparisons) of the LSA $cos(\theta)$ values suggest that our negative words were more cohesive as a category than both our neutral words [$t(126) = 7.92$, $p < .001$, $d = 1.40$] and our categorized words [$t(126) = 3.19$, $p < .01$, $d = .56$]. The LSA measure of similarity suggests that negative words had *more* within-pool similarity than the categorized words. This contrasts with our subjective ratings, where no difference was measurable between the negative and category words. Because our norming study was conducted with participants drawn from the same subject population as our cued recall study, we suggest that our subjective norms are the more precise measure of how semantically related our cued recall participants would perceive the paired words to be. Our categorized words were also more cohesive as a category than our neutral words [$t(126) = 3.59$, $p < .001$, $d = .63$].

Each problem remained in the center of the screen for 5000 ms. The participant was asked to type the correct answer during this fixed interval, after which the screen was cleared for 200 ms.

During cued recall, a probe word was presented along with a blank line either to the left or right of the word, depending on whether the target word had been presented as the first or second item in the pair, respectively. Participants were asked to recall the word that was paired with the probe word during the study phase, type their responses into the computer, and press the "Enter" key. Within each study set, half the cued recall probes for each pair type were in the forward direction and half were in the backward direction. Note, as the participant was unable to predict the testing direction for a given pair, this further prevented any potential encoding preferences for one part of the association (e.g., 'central/peripheral' trade-offs in attention).

Participants had a maximum of 15,000 ms to respond, after which the screen was cleared for 250 ms. If participants could not recall a target word for the probe word, they were instructed to type "PASS". Misspellings or variants of the correct word were scored as incorrect responses. Note that responses were also analyzed using a spell-checking search algorithm using the common UNIX spell-checking program aspell (Philips, 1990, 2000). All incorrect responses were processed by the algorithm and were marked as correct if the correct response was found in the list of possible corrections, as done in our previous study (Madan et al., 2010). Since analyses with responses both before and after spell-checking were not substantially different, we report only analyses using the strict spelling criterion for accuracy. Response time was logged both when the participant pressed the first key of their response, as well as after they pressed "Enter" to submit their response. However, response time measures yielded no additional information (e.g., no speed-accuracy trade-off) and as such will not be discussed further.

Following the eight study sets there was a final free recall task. Participants had five minutes to recall as many words as they could remember from the experiment. Participants were instructed to type in a word and press the "Enter" key. Once a participant pressed the "Enter" key, the screen cleared and the participant was allowed to type in another word. Repeated responses were only counted once. The task was implemented with the Python experimental library (pyEPL; Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

4. Results and discussion

All analyses are reported with Greenhouse-Geisser correction for non-sphericity where appropriate. Effects were considered significant based on an alpha level of 0.05.

4.1. Final free recall

We first asked whether our methods replicated the enhancement of item-memory due to arousal, by examining probability of final free recall (Fig. 1). Participants in the Negative group recalled more negative words than neutral words [$t(119) = 2.72$, $p < .01$, $d = .32$; Negative: mean = .22; Neutral: mean = .19; for CIs, see Fig. 1]. Participants in the Category group recalled more categorized words than neutral words [$t(112) = 6.17$, $p < .001$, $d = .58$; Categorized: mean = .23; Neutral: mean = .19]. However, there was no significant difference between recall rates for negative and categorized words [$t(231) = 1.26$, $p > .1$, $d = .16$]. Similar to previous studies, our results suggest that enhancement of item-memory due to arousal, comparing negative and neutral words, can be explained as an effect of semantic cohesiveness alone (cf. Buchanan et al., 2006; Talmi & Moscovitch, 2004). Between the two participant groups, no difference was observed in recall of the neutral words [$t(231) = 1.36$, $p > .1$, $d = .18$].

4.2. Cued recall accuracy—negative group

We conducted a TARGET TYPE[2] × ASSOCIATION TYPE[2] × TEST DIRECTION[2] repeated-measures ANOVA, with cued recall target accuracy as the dependent measure. The levels of TARGET TYPE were 'negative' and 'neutral'. ASSOCIATION TYPE was either 'pure' or 'mixed'. TEST DIRECTION was either 'forward' or 'backward'. 'Pure' pair accuracy included cued recall accuracy from the NN and nn pairs—pairs for which both the probe and target were of the same item-type (negative or neutral). 'Mixed' pair accuracy was the cued recall accuracy from the Nn or nN pairs—for which the probe and target are of different item-types. See Table 2 for a demonstration of how our experimental conditions mapped onto our ANOVA factors. This ANOVA design was also conducted on the Category group in an analogous manner.

Cued recall accuracy for the Negative group is plotted in Fig. 2a. We found a significant main effect of ASSOCIATION TYPE, where pure pairs were recalled with greater accuracy than mixed pairs $\left[F(1,119) = 15.65,\ p < .001,\ \eta_p^2 = .12\right]$. This main effect was qualified by the interaction between TARGET TYPE and ASSOCIATION TYPE [$F(1,119) = 43.95$, $p < .001$, $\eta_p^2 = .27$]. Simple effects analyses found that pure-pair negative words were recalled worse than pure-pair neutral words [$t(119) = 4.66$, $p < .001$, $d = .35$]. However, mixed-pair negative words were recalled better than mixed-pair neutral words [$t(119) = 4.74$, $p < .001$, $d = .35$]. TARGET TYPE had no significant main effect $\left[p > .1,\ \eta_p^2 < .001\right]$.

TEST DIRECTION had no significant main effect $\left[p > .1,\ \eta_p^2 = .001\right]$ and was not involved in any significant interactions [all η_p^2's <.02]. This means that test direction *per se* did not significantly affect performance. However, the remaining factors may tell us whether it matters which item was the probe or target, regardless of test direction (i.e., in the form of an interaction between ASSOCIATION TYPE and TARGET TYPE). The symmetry this implies in the pure pairs is a replication of numerous prior findings of equivalent forward and backward recall in homogeneous pairs (Kahana, 2002). Furthermore, the lack of significant effects of TEST DIRECTION means we found no evidence for asymmetric influences of arousal on recall,

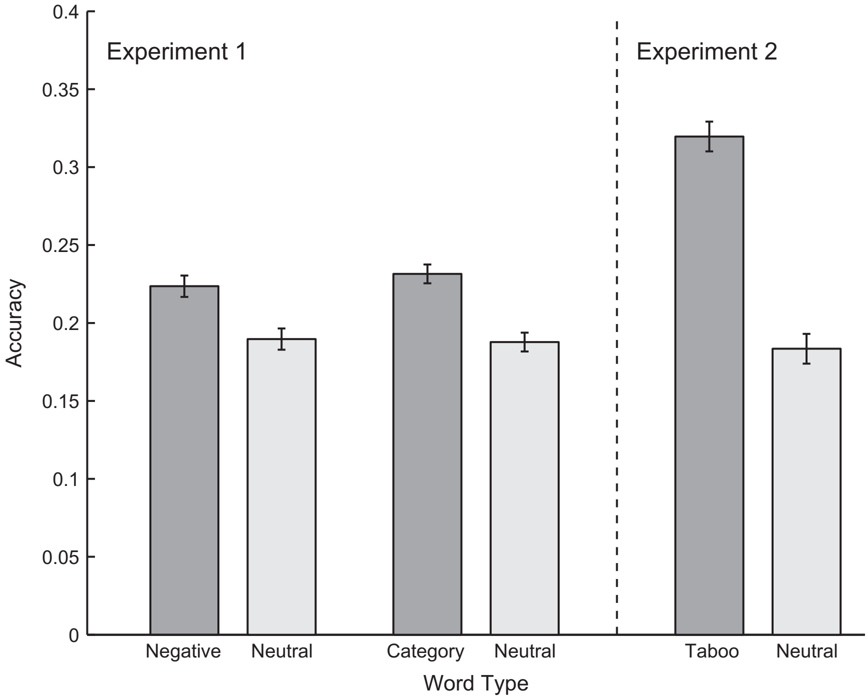

Fig. 1. Proportion of studied items recalled in final free recall, for both experiments. Error bars represent 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994).

implied by previous findings of temporally asymmetric effects of arousal on memory (e.g., Miu et al., 2005; Strange et al., 2003).

4.3. Cued recall accuracy—category group

Performance in this group is shown in Fig. 2b. We observed main effects of TARGET TYPE and ASSOCIATION TYPE. The main effect of TARGET TYPE found that categorized words were recalled better as targets than neutral words $\left[F(1,112) = 9.74,\ p < .01,\ \eta_p^2 = .08\right]$. The main effect of ASSOCIATION TYPE showed that accuracy in pure pairs was better than accuracy in mixed pairs $\left[F(1,112) = 45.38,\ p < .001,\ \eta_p^2 = .28\right]$. The interaction was not significant $\left[p > .1,\ \eta_p^2 < .001\right]$. Similar to the Negative group, TEST DIRECTION neither had a significant main effect $\left[p > .1,\ \eta_p^2 = .01\right]$ nor was not involved in any significant interactions [all η_p^2's <.01].

4.4. Model-based estimation of the arousal/category effects on item- and association-memory

To quantify the relative effects of arousal and category properties on item-memory versus association-memory we fit mean accuracy with a probabilistic "item-relationship" model (Madan et al., 2010). Our model is based on multiplication as it assumes that there are three separable processes involved in successfully recalling a target item in cued recall. Each of these processes has a probability of being completed successfully. By this logic, the probability that all three processes will be successful is the result of multiplying the probabilities from the three processes together. In the model, $Acc(\text{Pair Type}, \text{Test Direction})$, denotes accuracy for all eight possible pair type × test direction conditions, and is the product of probabilities representing (a) how effectively the probe item can be used to access the association in memory, (b) the strength of a representation of the association in memory, and (c) whether the target item will be produced correctly.

$$Acc(\text{Pair Type}, \text{Test Direction}) = P(Probe_i) \times P(Relat_j) \times P(Target_k)$$

Using the Negative group as an example, $P(Probe_i)$ denotes the probability of effectively using the probe item to access the association in memory, where, i = N,n. $P(Relat_j)$ denotes the probability corresponding to the strength of the association in memory, where j = NN, Nn, nN, nn. $P(Target_k)$ denotes the probability of retrieving the target item, where k = N, n. This results in the following system of equations:

$$Acc(\text{NN}, \text{Forward}) = P(Probe_\text{N}) \times P(Relat_\text{NN}) \times P(Target_\text{N})$$
$$Acc(\text{NN}, \text{Backward}) = P(Probe_\text{N}) \times P(Relat_\text{NN}) \times P(Target_\text{N})$$
$$Acc(\text{Nn}, \text{Forward}) = P(Probe_\text{N}) \times P(Relat_\text{Nn}) \times P(Target_\text{n})$$
$$Acc(\text{Nn}, \text{Backward}) = P(Probe_\text{n}) \times P(Relat_\text{Nn}) \times P(Target_\text{N})$$
$$Acc(\text{nN}, \text{Forward}) = P(Probe_\text{n}) \times P(Relat_\text{nN}) \times P(Target_\text{N})$$
$$Acc(\text{nN}, \text{Backward}) = P(Probe_\text{N}) \times P(Relat_\text{nN}) \times P(Target_\text{n})$$
$$Acc(\text{nn}, \text{Forward}) = P(Probe_\text{n}) \times P(Relat_\text{nn}) \times P(Target_\text{n})$$
$$Acc(\text{nn}, \text{Backward}) = P(Probe_\text{n}) \times P(Relat_\text{nn}) \times P(Target_\text{n})$$

Table 2
Factorial design of the pair types and test directions used in our study, using the Negative group as an example. Types of probe, relationship, and target are listed for all possible pair type × testing direction combinations.

Pair type	Testing direction	Probe	Relationship	Target
NN	Forward	N	NN (Pure)	N
	Backward	N	NN (Pure)	N
Nn	Forward	N	Nn (Mixed)	n
	Backward	n	Nn (Mixed)	N
nN	Forward	n	nN (Mixed)	N
	Backward	N	nN (Mixed)	n
nn	Forward	n	nn (Pure)	n
	Backward	n	nn (Pure)	n

The left-hand sides of these equations correspond to the conditions in our factorial design (Table 2). Because we are interested in the relative effects of stimulus properties on each of these three processes, we define the following ratio parameters which will be used as free parameters in the model fits:

$$p = \frac{P(Probe_{N})}{P(Probe_{n})}$$

$$r_1 = \frac{P(Relat_{NN})}{P(Relat_{Nn,nN})}$$

$$r_2 = \frac{P(Relat_{Nn,nN})}{P(Relat_{nn})}$$

$$t = \frac{P(Target_{N})}{P(Target_{n})}$$

If any ratio parameter were equal to the value 1, this would represent a null effect of arousal on the mechanism in the model. If the parameter were significantly greater than 1, this would indicate a significant advantage due to arousal in the respective memory process. For example, if $t > 1$, negative words were more retrievable target items than neutral words. Similarly, if $p > 1$, negative words would be more effectively accessed as probes than neutral words. If the ratio parameter were significantly less than 1 in the model fit, this would suggest that the respective memory process would be impaired by arousal.

Thus far, our item-relationship model is underdetermined (there are multiple ways to explain the data using various combinations of parameters). For this reason, we only used further-constrained model variants wherein a subset of the parameters p, r_1, r_2, and t was fixed to 1 and the remaining parameters were free to vary. After constraining the model, the system of equations could be solved algebraically for each participant and then parameter values and model fits were summarized across participants. We calculated *BIC* (Bayesian Information Criterion) as a measure of model-fitness that takes into account the number of free parameters. By convention, if the difference between two model fits, $\Delta BIC < 2$, neither of the models' fit to the data is significantly better—thus we report all scores as ΔBIC relative to the best-fitting model (Burnham & Anderson, 2002, 2004).

4.5. Interpretation of model parameters

The goal of the model-based analysis is to be able to distinguish effects of arousal on association-memory *per se*, versus effects of arousal that influence cued recall accuracy but would be better understood as effects due to item-memory. Next we explain the model and give concrete examples of cognitive processes that each of the model parameters could plausibly reflect.

The t parameter represents an influence on cued recall performance that depends *only* on characteristics of the target word. Note that during study, participants cannot know which of the two words of a given pair will later become the target in cued recall. For this reason, t reflects effects of word properties that influence item retrieval processes. As an example, in many models of recall (e.g., SAM; Raaijmakers & Shiffrin, 1981), one assumes that many candidate words are sampled (analogous to drawing a piece of paper from a hat). Some item properties might make them more likely to be sampled (e.g., a word with a high word frequency would be written on several pieces of paper in the hat).

Similarly, the p parameter represents influences on cued recall that depend *only* on the properties of the probe word. Again, because either of the words presented during study may later become the probe during cued recall, p reflects word properties that influence cued recall performance at test. For example, one type of word might be more easily identifiable due to unique or distinctive features. Criss et al. (2011) demonstrated that contextual variability, a measure of the number of different contexts a word can be used in, can primarily influence probe effectiveness. Specifically, low, as opposed to high, contextual variability words served as better probe items in cued recall. This significant probe effect is likely due, at least in part, to low contextual variability words having less pre-existing associations competing with new episodic associations learned in the experiment, than compared to high contextual variability words. As a result, low contextual variability words are more distinctive (by having less contextual variability) and are more effective as probe items, without enhancing memory for associations *per se*.

The remaining two parameters, r_1 and r_2, refer to memory for the association and represent influences on cued recall due to the properties of *both* paired items together, regardless of which one will become the probe and which will become the target. r_1 represents the ratio of recalling NN relative to Nn and Nn pairs. r_2 represents the ratio of recalling Nn and nN relative to nn pairs. Because these parameters refer only to the properties of the word pair, they are agnostic as to which item is the probe or target; thus, r_1 and r_2 can identify effects in cued recall that

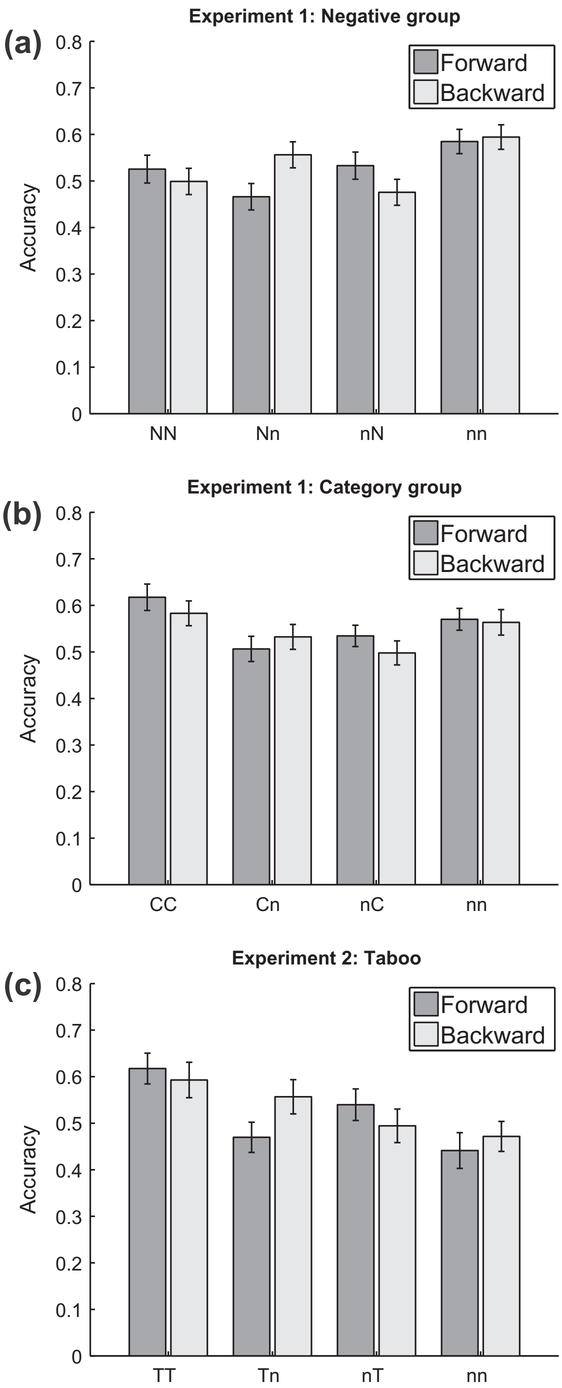

Fig. 2. Cued recall accuracy across pair types. Error bars represent 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994). (a) Accuracy for the Negative group of Experiment 1. Pair types: negative–negative (NN), negative–neutral (Nn), neutral–negative (nN), and neutral–neutral (nn). (b) Accuracy for the Category group of Experiment 1. Pair types: categorized–categorized (CC), categorized–neutral (Cn), neutral–categorized (nC), and neutral–neutral (nn). (c) Accuracy for the taboo manipulation of Experiment 2. Pair types: taboo–taboo (TT), taboo–neutral (Tn), neutral–taboo (nT), and neutral–neutral (nn).

relate to enhanced association-memory, independent of any memory enhancements or impairments due to the properties of the probe word or the target word. The resulting association-memory changes would only depend on the presence of the arousing word *within the respective pair* but not on properties of the arousing word by itself. If cued recall performance is a linear function of the item property being manipulated, with, hypothetically, recall being best for NN pairs, intermediate for mixed pairs (Nn and nN), and worst for nn pairs, we would expect $r_1 = r_2$. However, it is quite plausible that recall varies non-linearly as a function of item property. It is also possible that recall is a non-monotonic function of the item property. For instance, pure pairs may be better remembered than mixed pairs, due to associability or compatibility between the words that make up each pairs. We may then instead find that $r_1 = 1/r_2$. To illustrate this within our experiment, it is likely that pure pairs of words belonging to the same semantic category (Category group) would be better recalled than mixed pairs. As these relationships are usually unknown, we allow r_1 and r_2 to be fit separately and independently.

Note, because of associative symmetry (Kahana, 2002), these kinds of effects should influence forward and backward probes of a given pair type equivalently, i.e., a nN pair tested in the forward direction would be the same as a Nn pair tested in the backward direction. In both cases the probe word is negative (N), the association is mixed (Nn or nN), and the target word is neutral (n). Our ANOVA results confirmed that test direction was not involved in any main effects or interactions on cued recall, and therefore, test direction was not included in our models.

4.6. Approach to model selection

We fit six constrained models, with one or more parameters fixed to 1, to our behavioral data. In the first constrained model, we tested how much of the cued recall performance was accounted for by arousal effects on item retrieval. This model is called the 'Target-only' model, where only the t parameter was allowed to freely vary. In the second model we tested effects of pure item-memory in a 'Probe-only' model (only the p parameter was allowed to vary freely). This model tested how much of the cued recall performance was accounted for by arousal effects on the ability to effectively use the probe item. In the third constrained model, we tested a 'Relationship-only' model, where $p \equiv t \equiv 1$, leaving two free parameters, r_1 and r_2. Although the pure item-effect models (the 'Target-only' model and 'Probe-only' model) had fewer degrees of freedom, our fitness measure (ΔBIC) includes a penalty for degrees of freedom. Fig. 3 illustrates how these highly constrained models might explain our data. An enhancement of item retrieval due to arousal, with no effect on association-memory, would produce a pattern of behavior similar to Fig. 3a. An enhancement of the efficacy of probe items due to arousal would appear similar to the pattern in Fig. 3b. If arousal enhanced association-memory without any effect on item-memory, results would appear similar to Fig. 3c. Finally, an impairment of association-memory would appear similar to Fig. 3d.

The fourth constrained model contained the two item-memory parameters: p and t (two free parameters; 'Probe and Target' model). The final two constrained models involved the relationship parameters and either one of the item parameters, resulting in free parameters r_1 and r_2 in addition to either p or t (three free parameters in total; 'Relationship and Probe' and 'Relationship and Target' models, respectively).

To compute 95% confidence intervals of the model parameters obtained for each model variant, we applied a bootstrap, randomly sampling participants with replacement (an improvement on the method we used in Madan et al., 2010). Each of 10,000 such random samples contained the same amount of data. We then fit each model variant to each resampled data set and report the corresponding 2.5 and 97.5 percentile values as the limits of a two-tailed 95% confidence interval.

After determining the fits for each of the six constrained models, multiple models can sometimes fit the data equally well. In the final stage of model selection, we examined the confidence intervals for each of the parameters in the best-fitting models and drew on converging evidence from the final free recall data and previous research. The same approach was used to fit the data in the Category group and in Experiment 2.

4.6.1. Negative group

The two models involving the relationship and one item parameter ('Relationship and Probe' and 'Relationship and Target' models) provided the best fit by far, suggesting that the effect of arousal on cued recall performance influences both item- and association-memory (Table 3; Fig. 4f). Note that due to model mimicry, both 'Relationship and Probe' and 'Relationship and Target' models produce identical fits to the empirical data, but they do so with different parameter values. Corroborating evidence from our own data and previous research was then used to drive our ultimate model selection. In final free recall, we found better memory for negative words than for neutral words. This suggested that cued recall accuracy should incorporate enhancement of item-memory ($t > 1$). Additionally many studies have found an enhancement of item-memory for arousing words (Anderson, Wais, & Gabrieli, 2006; Phelps, 2004; Rubin & Friendly, 1986). However, the 'Relationship and Probe' model cannot account for any differences in target retrievability due to arousal (because t was fixed to 1), and instead suggests that arousal *impairs* probe effectiveness, a result that is not supported by existing research into emotion and memory. Thus, we selected the 'Relationship and Target' model as the more plausible model. The best-fitting parameters in this model suggest that arousing items damage association learning (r_1 and $r_2 < 1$), that arousal also enhanced target retrieval ($t > 1$), but that arousal had no influence on probe effectiveness ($p = 1$).

4.6.2. Category group

We found three equivalently best-fitting models: the 'Relationship-only' model, the 'Relationship and Probe' model, and the 'Relationship and Target' model (Table 3; Fig. 5). The only difference between categorized words and neutral words was the relationship between the words, based on stronger pre-experimental associations between some of the items (the categorized words) than others (the neutral words). Thus, this relationship should influence cued recall performance, which is reflected by the fact that all best-fitting models included the relationship parameters. Based on converging evidence from the final free recall task, where more categorized words were recalled than neutral words, cued recall performance should also be influenced by enhanced item-retrievability due to semantic cohesiveness (t). Again, the 'Relationship and Probe' model cannot account for any differences in target retrievability due to semantic cohesiveness, and also suggests no significant effect of semantic cohesiveness on probe effectiveness (p). Thus, strongest support was present for the 'Relationship and Target' model.

4.6.3. Between-group comparisons

We included the Category group as a control to test whether arousal effects on memory could be explained by a closer semantic cohesiveness among negative words than neutral words. The r_1 parameter being *below* zero in the Negative group implies difficulties in association-memory when two negative words are linked to each other as opposed to one negative and one neutral word. r_1 being *above* zero in the Category group in turn suggests that association-memory of two words from a semantically cohesive category helps rather than hinders association-memory. That is, semantic relatedness alone cannot account for the reduction in association-memory for negative words.

Independent model selection for the Negative and Category groups suggested the same best-fitting model—the 'Relationship and Target' model. Thus, here we compare the best-fitting model parameters for each group directly. To ask whether best-fitting model parameters were different between the two groups, we compared each parameter between groups with t-tests (see Table 3 for parameter fits). Parameter values were log-transformed to satisfy the normality assumption of the t-test. r_2 and t were not significantly different between groups [$p > .1$]. This lack of significant difference of t between Negative and Category groups supports the findings of Talmi and Moscovitch (2004), who found that the enhancement of item-memory due to arousal for single words could be explained solely by semantic cohesiveness. Importantly, r_1 was significantly different between conditions [$t(228) = 4.28$, $p < .001$]. Thus, between-group differences in r_1 suggest that arousal-induced impairments of association-memory observed in our data cannot be explained by semantic cohesiveness alone. In contrast, the finding that r_2 was below 1 in both the Negative and Category group provides evidence that shared aspects of both manipulations (i.e., semantic cohesiveness) was negatively associated with association-memory.

One caveat to using data from our Category group as a control for our Negative group, is that participants in our Category group were given words from four distinct semantic categories ('driving', 'school', 'house', 'business'), whereas participants in our Negative group were only given words from a single set that could be functioning like a semantic category (i.e., 'negative'). Thus, it is possible that Category group has reduced category effects due to the use of multiple categories. To account for this we re-analyzed the data using only participants' first study

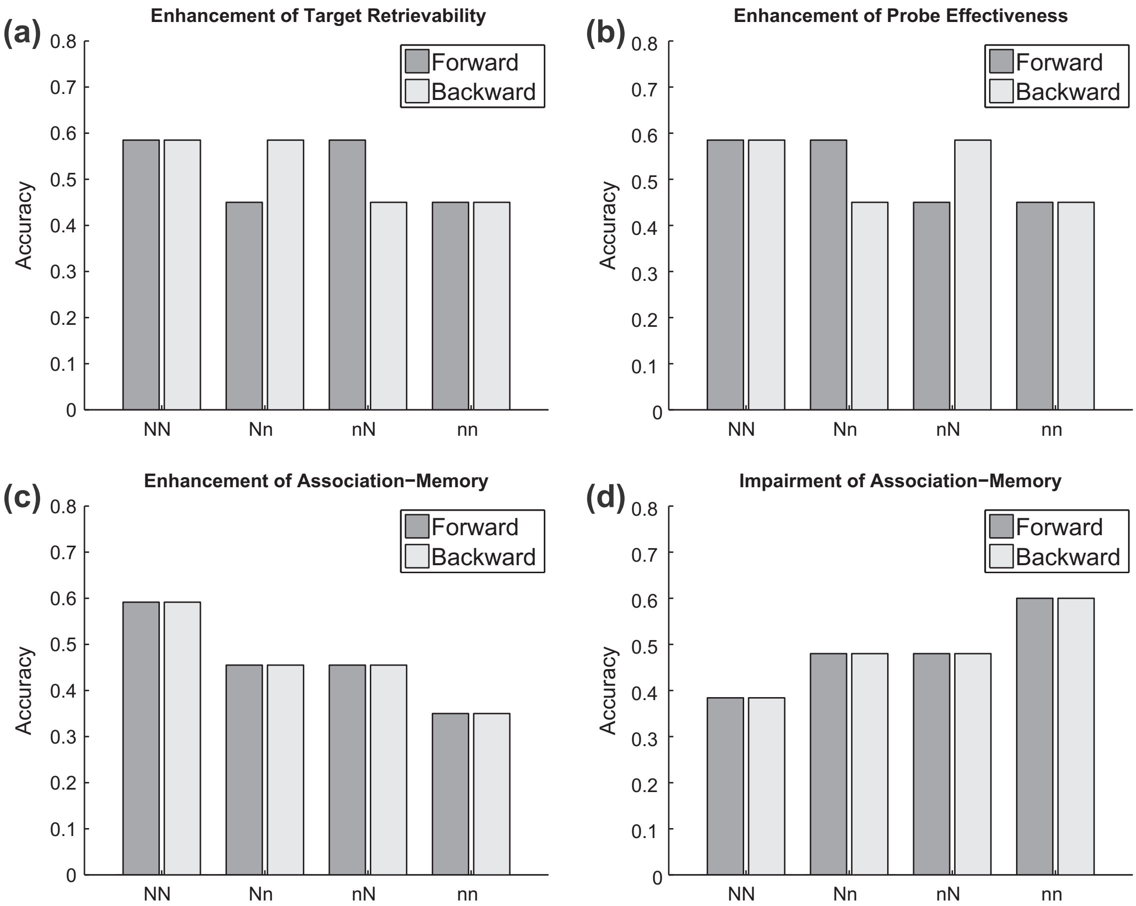

Fig. 3. Simulations of memory effects on mean accuracy. Pair types: negative–negative (NN), negative–neutral (Nn), neutral–negative (nN), and neutral–neutral (nn). (a) Simulated effect of enhanced item retrieval ($t = 1.30$). (b) Simulated effect of enhanced probe effectiveness ($p = 1.30$). (c) Simulated effect of an enhancement of association-memory ($r_1 = 1.30$, $r_2 = 1.30$). (d) Simulated effect of an impairment of association-memory ($r_1 = 0.80$, $r_2 = 0.80$). See main text, modeling section, for an explanation of the model parameters.

set and fit our model to this data. At this point, participants have only been exposed to one semantic category and possible confounds based on our use of multiple categories is not yet an issue. We then fit the 'Relationship and Target' model (the best-fitting model for the complete data set), to this reduced data set. In the Category group, we found the best-fitting model parameters to qualitatively agree with our full data set ($r_1 = [0.99, 1.39]$, $r_2 = [0.69, 0.98]$, $t = [0.89, 1.29]$). As in the full data set, r_1 was (almost significantly) greater than 1, and r_2 was (significantly) less than 1. In the Negative group, the model parameters generally had larger variability, but qualitatively were in agreement with the model fits to the full data set ($r_1 = [0.84, 1.16]$, $r_2 = [0.75, 1.03]$, and $t = [0.99, 1.36]$). Nonetheless, even though the parameter ranges are large when the model is only applied to the first study set, it is evident that r_1 and r_2 move in the opposite direction in the Category group, where r_1 becomes greater than 1, while r_2 is less than 1. In contrast, in the Negative group, both r_1 and r_2 are largely consistent. Importantly, r_2 appears to be similar between the Negative and Category groups, even in the first study set, whereas parameter estimates for r_1 appear to diverge qualitatively, with the range for r_1 for the Category group being larger in magnitude than the range for the Negative group. Additionally, even if our model parameter estimates are reduced due to the use of multiple categories, in our full data set (Table 3) we found that r_1 was significantly greater than 1 in our Category group, while r_1 was significantly less than 1 in our Negative group. Thus, it appears unlikely that the r_1 parameter would be qualitatively different if we had used just one instead of four semantic categories.

4.7. Discussion

In Experiment 1, we found that although arousal enhances item-memory, association-memory can simultaneously be impaired, supporting our *associative impairment hypothesis*. This occurred even when association-learning was the primary objective of the task and full attention could be devoted to both parts of the association due to sequential presentation. Furthermore, we observed that arousal influences memory for the association itself, ruling out our *item-memory hypothesis*: That is, the cued recall reduction

Table 3
Model fits for both Experiment 1 and Experiment 2. All model variants are shown, with the exception of the full model (as it is underdetermined by the data). All free parameter fits are presented as 95% confidence intervals. Note that the 'Relationship and Target' and the 'Relationship and Probe' models algebraically produce identical fits due to model mimicry, although their best-fitting parameters are not equivalent.

	ΔBIC	p	r_1	r_2	t
Experiment 1: Negative group					
Target-only	18.1	1	1	1	[0.96, 1.04]
Probe-only	7.92	[0.83, 0.90]	1	1	1
Relationship-only	12.6	1	[0.96, 1.07]	[0.82, 0.90]	1
Probe and Target	14.8	[0.83, 0.90]	1	1	[0.96, 1.04]
Relationship and Target[a]	0	1	[0.88, 1.00]	[0.75, 0.84]	[1.09, 1.22]
Relationship and Probe	0	[0.81, 0.92]	[1.02, 1.16]	[0.88, 0.97]	1
Experiment 1: Category group					
Target-only	5.15	1	1	1	[0.92, 1.18]
Probe-only	6.28	[0.88, 1.11]	1	1	1
Relationship-only	0.0154	1	[1.04, 1.40]	[0.75, 0.99]	1
Probe and Target	15.2	[0.88, 1.11]	1	1	[0.92, 1.18]
Relationship and Target[a]	0	1	[1.01, 1.38]	[0.69, 0.97]	[0.88, 1.29]
Relationship and Probe	0	[0.77, 1.13]	[1.04, 1.51]	[0.75, 1.04]	1
Experiment 2: Taboo					
Target-only	8.23	1	1	1	[1.16, 1.31]
Probe-only	20.9	[1.03, 1.15]	1	1	1
Relationship-only	11.4	1	[1.10, 1.25]	[1.04, 1.23]	1
Probe and Target[a]	0	[1.03, 1.15]	1	1	[1.16, 1.31]
Relationship and Target	8.48	1	[1.03, 1.19]	[0.97, 1.15]	[1.06, 1.22]
Relationship and Probe	8.48	[0.82, 0.94]	[1.16, 1.36]	[1.11, 1.31]	1

[a] The best-fitting models according to our model-fitness measure (ΔBIC) and additional converging evidence.

in arousing pairs was not due solely to item-memory effects. While our results suggest that the enhancement of item-memory caused by our negative words can be accounted for by semantic cohesiveness (as suggested by Talmi & Moscovitch, 2004), the results of our modeling analyses suggest that the impairing effect of arousal on association-memory was not simply a result of closer semantic cohesiveness between negative words. In particular, we found evidence of an enhancement of association-memory due to semantic cohesiveness ($r_1 > 1$) in our Category group, while we found an impairment of association-memory in our Negative group ($r_1 < 1$).

Interestingly, Siddiqui and Unsworth (2011) recently came to a different conclusion based on two free recall studies. In these studies, participants learned mixed lists of positive, negative, and neutral words. In free recall, both types of emotional words were retrieved earlier than neutral words. There also was a clustering in recall of same-valence emotional words (Experiments 1 and 2), especially when the encoding task oriented participants towards the valence of the words through pleasantness ratings (Experiment 2). According to the authors, these results could imply that emotion enhances the probability of an item being sampled (an item's relative strength; indexed by earlier retrieval) *as well as* its inter-item associations (demonstrated by the clustering of negative and positive words in recall). Although our results would concur with their first suggestion, i.e., an enhancement of item-memory through emotional arousal, we observed a reduction rather than strengthening in association-memory due to emotional arousal. Critically, free recall does not demand that any particular inter-item associations be formed, whereas paired-associate learning does. Thus, when asked to do so, participants in our study were less able to form associations between the two specific words of a given pair, if the pair involved a negative word, compared to two neutral words. Conversely, when the task does not demand memory for specific pairings as in free recall, emotional words may indeed cue memory for other emotional words. However, semantic similarity between emotional and neutral words was not equated in Siddiqui and Unsworth (2011) and the clustering pattern in recall appeared to be valence-specific, suggesting that a higher semantic cohesiveness of negative or positive words compared to neutral words likely contributed to their observation of valence-based clustering in free recall. Since we found here that negative words impaired association-memory whereas neutral but equally semantically cohesive words did not, we can rule out that negative emotion mimicked effects of semantic similarity on association-memory in paired-associate learning.

5. Experiment 2

One possible reason we may have observed an impairment of association-memory due to arousal is that our moderately arousing, negative words may not have been arousing enough to cause any potential enhancement of association-memory. Taboo words more reliably evoke emotional arousal than negative words (e.g., MacKay et al., 2004; Siegrist, 1995). Importantly, Guillet and Arndt (2009) only found an enhancement of association-memory when using taboo words as probe items, but found no significant effect on association-memory when using negative words. Thus, to further test the influence of arousal on association-memory, in Experiment 2 we used taboo words instead of the moderately arousing negative words used in Experiment 1. As in Experiment 1, we began with three alternative hypotheses:

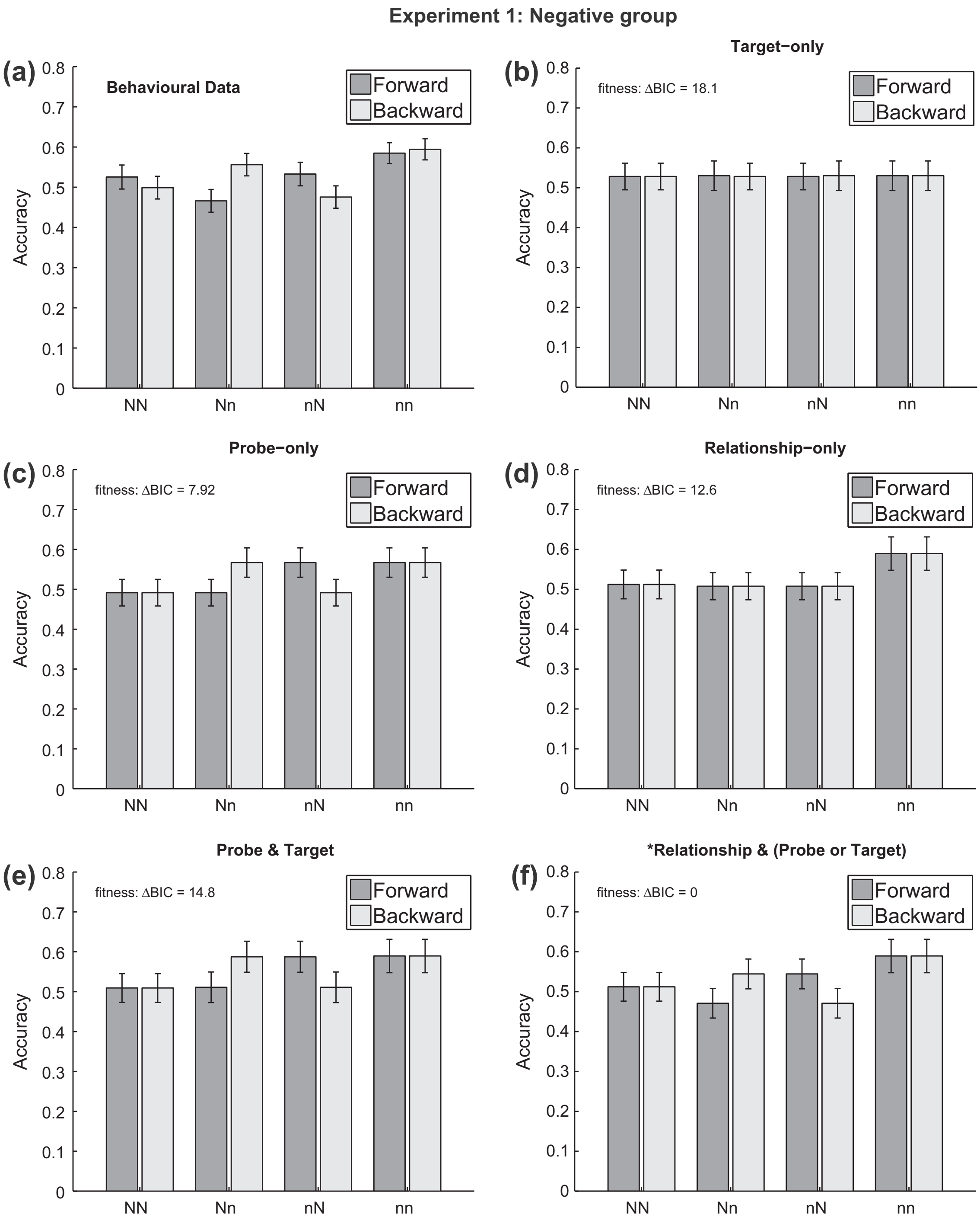

Fig. 4. Modeling of cued recall for the Negative group. Pair types: negative–negative (NN), negative–neutral (Nn), neutral–negative (nN), and neutral–neutral (nn). * Denotes the best-fitting models according to our model-fitness measure (ΔBIC) and additional converging evidence. Error bars represent 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994). (a) Behavioral data, repeated from Fig. 2a to clarify comparisons between data and model variants. (b) 'Target-only' model (t = [0.96, 1.04]). (c) 'Probe-only' model (p = [0.83, 0.90]). (d) 'Relationship-only' model (r_1 = [0.96, 1.07], r_2 = [0.82, 0.90]). (e) 'Probe and Target' model (p = [0.83, 0.91], t = [0.96, 1.04]). (f) Relationship and (Probe or Target) models: 'Relationship and Target' model (r_1 = [0.88, 1.00], r_2 = [0.75, 0.84], t = [1.09, 1.22]) and 'Relationship and Probe' model (p = [0.81, 0.92], r_1 = [1.02, 1.16], r_2 = [0.88, 0.97]).

5.1. Association-memory enhancement hypothesis

It is possible that the negative words used in Experiment 1 were not sufficiently arousing at a physiological level, and that an enhancement of association-memory may emerge, overcoming the association-memory impairment we observed in Experiment 1. For instance, Guillet and Arndt (2009) found a reduction in cued recall accuracy,

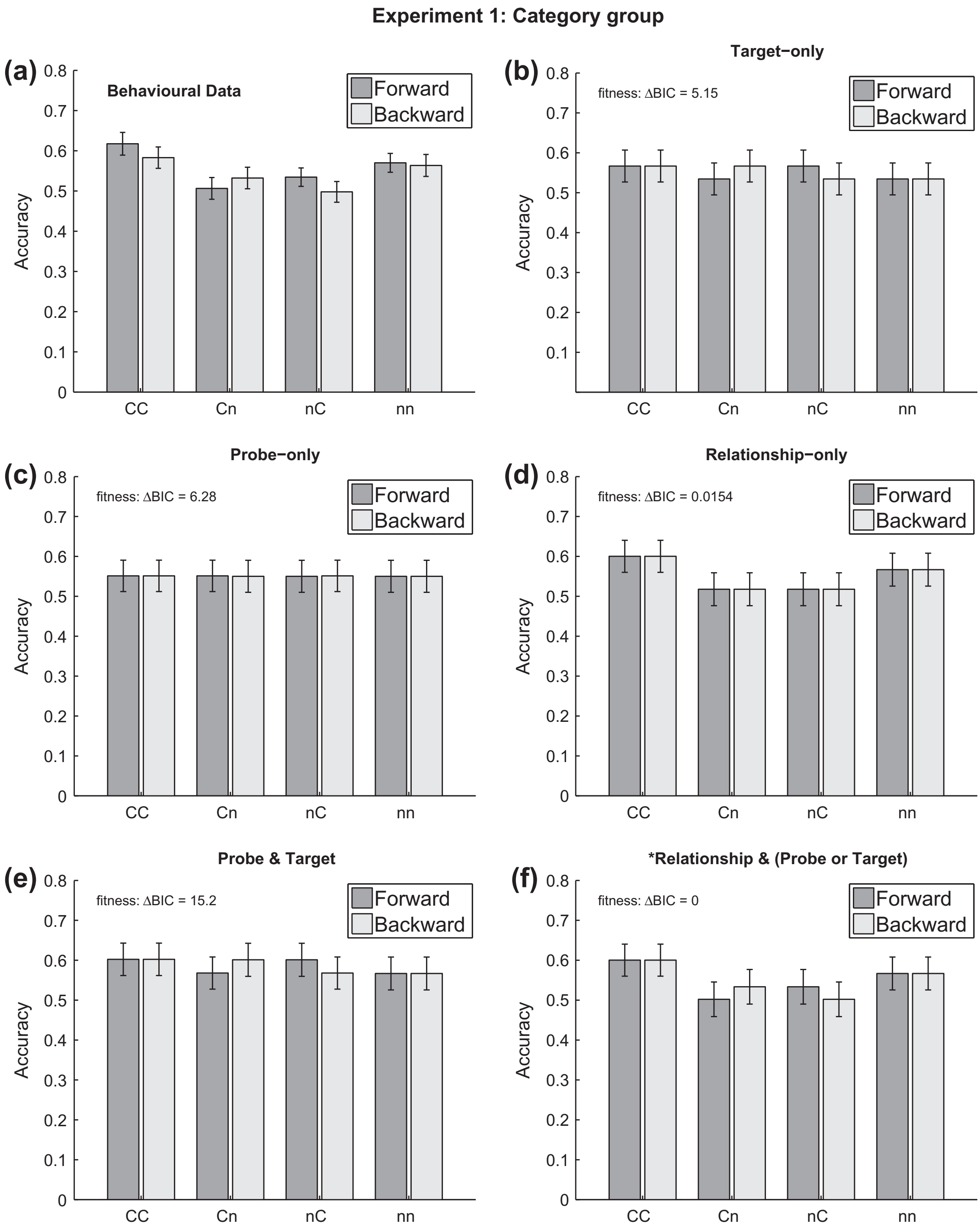

Fig. 5. Modeling of cued recall for the Category group. Pair types: categorized–categorized (CC), categorized–neutral (Cn), neutral–categorized (nC), and neutral–neutral (nn). * Denotes the best-fitting models according to our model-fitness measure (ΔBIC) and additional converging evidence. Error bars represent 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994). (a) Behavioral data, repeated from Fig. 2b to clarify comparisons between data and model variants. (b) 'Target-only' model ($t = [0.92, 1.18]$). (c) 'Probe-only' model ($p = [0.88, 1.11]$). (d) 'Relationship-only' model ($r_1 = [1.04, 1.40]$, $r_2 = [0.75, 0.99]$). (e) 'Probe and Target' model ($p = [0.88, 1.11]$, $t = [0.92, 1.15]$). (f) Relationship and (Probe or Target) models: 'Relationship and Target' model ($r_1 = [1.01, 1.38]$, $r_2 = [0.69, 0.97]$, $t = [0.88, 1.29]$) and 'Relationship and Probe' model ($p = [0.77, 1.13]$, $r_1 = [1.04, 1.51]$, $r_2 = [0.75, 1.04]$).

though not statistically significant, when using negative words as probe items relative to using neutral words. However, when using highly arousing taboo words as probe items they found enhanced cued recall performance.

Providing physiological evidence for possible diverging effects of arousal, Buchanan et al. (2006) conducted a free recall study with a mixed list composed of taboo, negative, categorized, and neutral words, while simultaneously recording participants' heart rate. Buchanan et al. (2006) observed that participant's heart rate accelerated when taboo words were presented, but decelerated with negative words, pointing to potentially diverging effects of taboo words (possibly a defensive response) versus negative words (an orienting response; Hare, 1973). Hadley and MacKay's (2006) priority-binding hypothesis suggesting better binding of highly arousing items to their context also mainly draws evidence from studies comparing taboo words with neutral words. Therefore, here we predict that memory for the association will be greater when the pair includes a taboo word than when a pair includes only neutral words: both r_1 and $r_2 > 1$.

5.2. Association-memory impairment hypothesis

In Experiment 1 we observed an impairment of association-memory due to arousal. If arousal truly enhances item-memory while simultaneously impairing association-memory, effects in Experiment 2 should be similar to those of Experiment 1, but larger in magnitude due to the higher arousing nature of taboo words. Therefore, here we predict that memory for the association will be reduced when the pair includes one or more taboo words than when a pair includes only neutral words: both r_1 and $r_2 < 1$.

5.3. Item-memory hypothesis

Many studies have shown that taboo words provoke an even greater enhancement in free recall due to arousal than moderately arousing words (e.g., Buchanan et al., 2006; Jay, Caldwell-Harris, & King, 2008; Kensinger & Corkin, 2003; Schmidt & Saari, 2007). Additionally, while taboo words are more arousing than other emotional words, they also contain a taboo-specific word property (e.g., "offensiveness") that make them more distinct. Even by using a higher arousal level as a manipulation, it is therefore still possible, unlike in the Negative group of Experiment 1, that memory for associations is unaffected. This would be reflected by both r_1 and $r_2 = 1$ in our model. In that case, cued recall performance could be driven primarily by taboo-specific word properties enhancing the item-strength for the taboo words. Therefore, here we predict both p and $t > 1$.

6. Methods

6.1. Participants

A total of 68 undergraduate students (mean age ± *sd* = 18.90 ± 3.02; 23 males and 45 females) participated in our study for partial credit in an introductory psychology course at the University of Alberta. Data from two of these participants was lost due to machine error. All participants were required to have learned English before the age of six and to be comfortable typing. Participants gave written informed consent prior to beginning the study, which was approved University of Alberta Research Ethics Board. None of the participants from Experiment 1 participated in Experiment 2.

6.2. Materials

A subset of 64 taboo words from Anderson (2005) was chosen, non-overlapping with any words from Experiment 1 (see Appendix B). An additional 64 neutral words were derived from Anderson (2005) and matched for length and word frequency using MCWord (Medler & Binder, 2005). (We used MCWord since the MRC Psycholinguistic Database did not have ratings for our taboo words.)

When initially choosing the words, Anderson (2005) was careful to match his taboo and neutral stimuli, allowing us not to have to worry about these issues ourselves. Nonetheless, when we, post hoc, used LSA to compare our taboo and neutral words, they were similarly cohesive[2] [$t(126) = 1.44$, $p > .1$, $d = .26$], which supports our omission of another category control group in Experiment 2. Despite the similarity one would expect between pairs of positive items, and pairs of negative items, the average similarity within the taboo pool was not different than that for the neutral pool. Thus, *within* Experiment 2, similarity is unlikely to explain a memory effect of taboo words compared to neutral words.

6.3. Procedure

Experiment 2 was identical to the Negative group in Experiment 1, with taboo words replacing negative words; thus, pair types were TT/Tn/nT/nn.

At the end of the experiment, participants were asked to rate all of the words first for arousal and then for valence. In the arousal and valence rating tasks, words were presented on the screen along with a 5-point version of the respective SAM diagram (Bradley & Lang, 1994). In the arousal rating task, '1' corresponded to excited and '5' corresponded to calm. In the valence rating task, '1' corresponded to pleasant and '5' corresponded to unpleasant.

Presentation order of words was randomized in each rating task. However, the order of judgements for each word was not randomized. Specifically, participants made all of the arousal judgements before all of the valence judgements, to avoid further attenuating participants' arousal judgements to due habituation from multiple exposures of the word.

7. Results and discussion

7.1. Arousal and valence ratings

Mean arousal and valence ratings for each word were calculated by averaging across participants and are reported

[2] We again calculated the LSA $cos(\theta)$ similarity for our word pools. LSA $cos(\theta)$ within each of our word pools (mean ± *sd*) are as follows: taboo (.113 ± .151) and neutral (.085 ± .104). We found that our taboo and neutral word pools were not differently cohesive [$t(124) = 1.44$, $p > .1$, $d = .26$]. As these two word pools were not significantly different in measures of cohesiveness, we did not also include a category-control group in Experiment 2. Note that two taboo words were not found in the LSA database and were excluded from these LSA calculations.

Table 4
Arousal and valence ratings are the subjective ratings according to the rating task used in Experiment 2. Frequency ratings were obtained from MCWord (Medler & Binder, 2005). Mean ratings are shown with standard deviation in parentheses. Means in a row with the same superscript are not significantly different at $p < .05$.

	Taboo	Neutral
Arousal	3.52 (0.73)[a]	4.58 (0.45)[b]
Valence	3.26 (0.45)[a]	2.86 (0.39)[b]
Frequency	14.6 (22.18)[c]	15.7 (16.25)[c]
Length	5.67 (1.53)[c]	5.42 (1.04)[c]

in Table 4. Taboo words were more arousing than neutral words [$t(65) = 13.85$, $p < .001$, $d = 1.76$]. Taboo words were more unpleasant than neutral words [$t(65) = 5.20$, $p < .001$, $d = .97$].

7.2. Final free recall

Participants recalled more taboo words than neutral words [$t(65) = 13.97$, $p < .001$, $d = 1.67$; Taboo: mean = .32; Neutral: mean = .18; for CIs, see Fig. 1]. Comparing the free-recall advantages of all three words types (negative, categorized, and taboo) over random neutral words across experiments, we found small [Negative group: $d = .32$] to moderate [Category group: $d = .58$] effect sizes in Experiment 1 but a large effect size in Experiment 2 [Taboo: $d = 1.67$]. This suggests a substantially larger free-recall advantage for taboo over neutral words than what we had observed for the negative words in Experiment 1. The enhancement of memory due to arousal was more pronounced, suggesting that taboo words had enhanced item-memory, over and above that which could be explained by semantic cohesiveness alone (see also Buchanan et al., 2006).

7.3. Cued recall accuracy

As in Experiment 1, a TARGET TYPE[2] × ASSOCIATION TYPE[2] × TEST DIRECTION[2] repeated-measures ANOVA was conducted on cued recall accuracy (Fig. 2c). TARGET TYPE had a significant main effect [$F(1,65) = 61.57$, $p < .001$, $\eta_p^2 = .49$], with better recall for taboo words than for neutral words. The interaction TARGET TYPE × ASSOCIATION TYPE was also significant [$F(1,65) = 8.61$, $p < .01$, $\eta_p^2 = .12$]. Taboo words from pure pairs were recalled better than taboo words from mixed pairs [$t(65) = 2.91$, $p < .01$, $d = .26$]. No significant difference was found in neutral word recall between pure and mixed pairs [$p > .1$, $d = .10$]. ASSOCIATION TYPE had no significant main effect $\left[p > .1,\ \eta_p^2 = .02\right]$. Again, TEST DIRECTION neither had a significant main effect $\left[p > .1,\ \eta_p^2 = .01\right]$, nor part of any significant interactions [all η_p^2's <.03].

In the Negative group of Experiment 1, the pure negative pairs were recalled significantly worse than the pure neutral pairs (NN < nn). However, in Experiment 2 the pure taboo pairs were recalled significantly *better* than the pure neutral pairs (TT > nn). This difference between experiments may be because association-memory now benefited from the higher arousal of the taboo words, but our model-based analysis will test this possibility directly.

7.4. Model-based estimation of the taboo effects on item- and association-memory

When fitting our model variants to the behavioral data (Table 3), the 'Probe and Target' model was clearly the best-fitting model (Fig. 6). This suggests that taboo words act as better probes *and* targets than neutral words ($p > 1$ and $t > 1$; see Table 3). This is in agreement with taboo words being recalled more than neutral words in free recall.

Although the model fits produced by the 'Relationship and Probe' and 'Relationship and Target' models appear to fit the behavioral data nearly as well as the 'Probe and Target' model, they led to significantly higher ΔBIC values, making model selection straightforward. In the 'Probe and Target' model, the manipulation (i.e., taboo) has no effect on a participant's ability to learn associations between items (represented by parameters r_1 and r_2 in the model). However, in this model, the manipulation influences a participant's ability to effectively use an item as a probe (parameter p) and retrieve items as targets (parameter t). If a taboo-specific word property enhances both p and t, we would find better recall for TT pairs than for nn pairs, as TT pairs both have taboo items as probes and targets. The clear success of the 'Probe and Target' model over all other models suggests that tabooness does not enhance or impair the relationship between the constituent items ($r_1 = r_2 = 1$). Thus, in increasing the arousal-level of our stimuli (through the use of taboo words), we increased their item-retrievability but did not observe the impairment of association-memory due to arousal we found in Experiment 1. Importantly, however, we neither observed enhancement nor impairment in the association-memory parameters.

7.5. Further model-based analyses

We considered potential limitations of our findings by re-analyzing our data in two ways. First, we wondered whether even our taboo stimuli were not sufficiently arousing; perhaps even with highly arousing stimuli, an underlying enhancement effect on association-memory due to arousal would emerge, similar to findings with pictures (Smith, Henson, Rugg, & Dolan, 2005). Based on the arousal ratings of our 64 taboo words obtained at the end of Experiment 2, we retained only the 32 most arousing words. Models were re-fit to this reduced dataset. Three models fit equally well: a 'Target-only' model, and two models now involving the association-memory parameters r_1 and r_2 ('Relationship and Probe', and 'Relationship and Target'). However, in the two latter models, r_1 and r_2 were not different from 1 based on 95% confidence intervals, and t in was still greater than 1 in the 'Target-only' and 'Relationship and Target' models. The outcome of model-selection on this reduced, higher-arousing data set is that high-arousing stimuli function better as targets, but even with a stronger manipulation of arousal, we still did not observe an enhancement of association-memory. Taboo words, of course, are still just an approximation of emotional arousal as it would be encountered in real-life situations and we cannot rule out the possibility that even more high-arousing materials

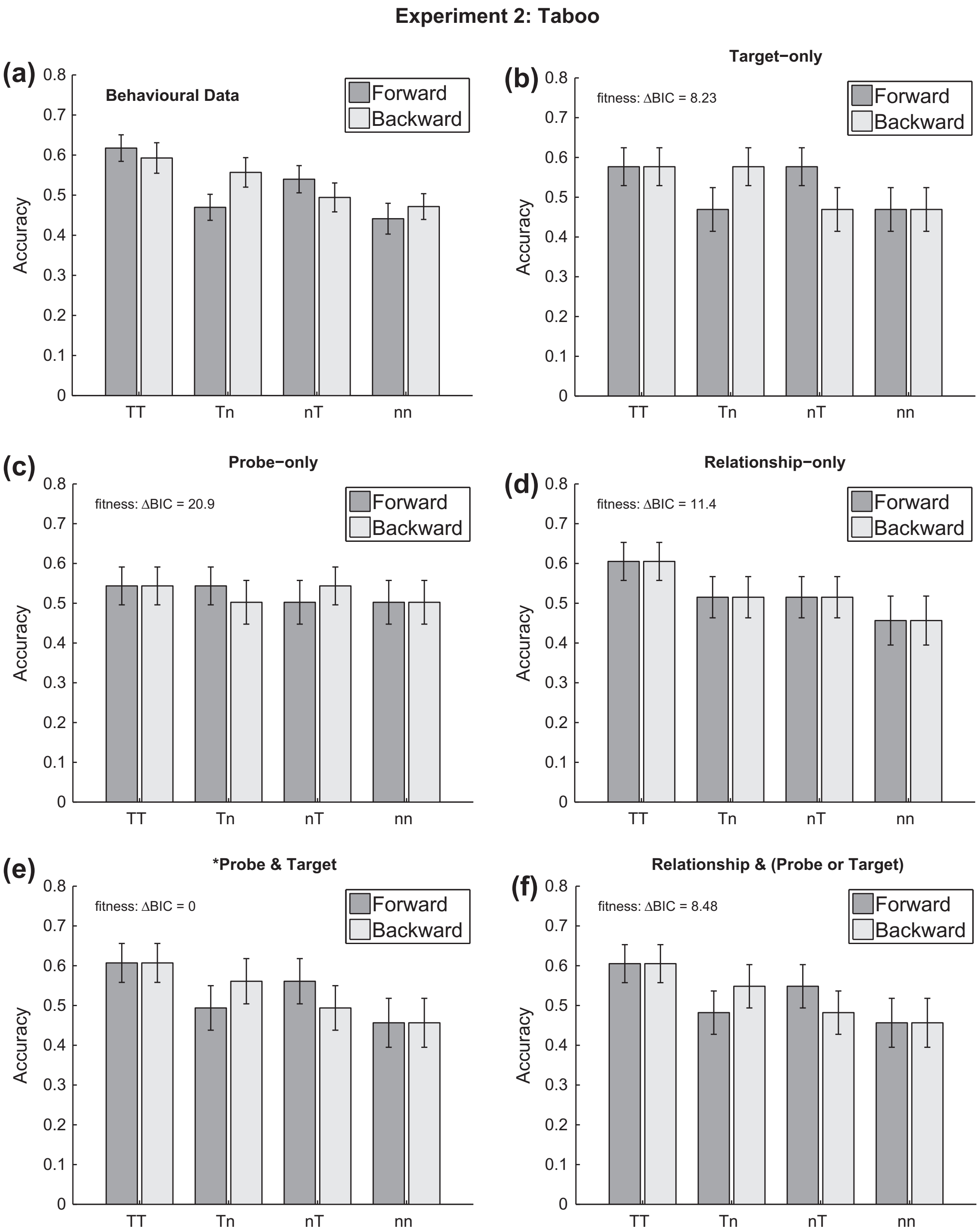

Fig. 6. Modeling of cued recall for the taboo manipulation. Pair types: taboo–taboo (TT), taboo–neutral (Tn), neutral–taboo (nT), and neutral–neutral (nn). * denotes the best-fitting models according to our model-fitness measure (Δ *BIC*) and additional converging evidence. Error bars represent 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994). (a) Behavioral data, repeated from Fig. 2c to clarify comparisons between data and model variants. (b) 'Target-only' model ($t = [1.16, 1.31]$). (c) 'Probe-only' model ($p = [1.03, 1.15]$). (d) 'Relationship-only' model ($r_1 = [1.10, 1.25]$, $r_2 = [1.04, 1.23]$). (e) 'Probe and Target' model ($p = [1.03, 1.15]$, $t = [1.16, 1.31]$). (f) Relationship and (Probe or Target) models: 'Relationship and Target' model ($r_1 = [1.03, 1.19]$, $r_2 = [0.97, 1.15]$, $t = [1.06, 1.22]$) and 'Relationship and Probe' model ($p = [0.81, 0.94]$, $r_1 = [1.16, 1.36]$, $r_2 = [1.11, 1.31]$).

could still enhance association-memory. However, based on the current results, we have no reason to expect that highly arousing events would ever increase association-memory *per se*. How can we integrate our findings with previous studies that reported ostensible enhancements of association memory by arousal? We suggest that alternate explanations based on item-memory or factors other than association-memory *per se* may have accounted for such findings. To illustrate with an example, Smith et al. (2005) observed better source-memory for neutral objects encoded in negative and positive scene contexts compared to neutral scene contexts. Hence, one could conclude that the arousing encoding context enhanced association-memory for object and context. However, accurate source judgements could have been due to (a) recollection of the arousing scenes by themselves (item-memory; analogous to our model parameter, t), (b) retrieval of the emotional valence of the encoding context without actual recollection of the context itself (reflecting a difference in accessibility due to arousal), or (c) actual retrieval of the context–object associations. Our findings suggest that both options (a) and (b) are quite plausible and the enhanced retrieval was not due to enhancement of association-memory itself.

Secondly, we used only negative words in Experiment 1. In contrast, our taboo words in Experiment 2 were on average negative, but many in Experiment 2 were positive both in the group and in individual participants (e.g., sex-related taboo words like ORGASM, SEX, CARESS, CLIMAX, SENSUAL). Thus, it is possible that the impairment of association-memory in Experiment 1 was caused by valence rather than arousal effects. To address this possible confound, we re-analyzed the data from Experiment 2 by retaining only the 32 most negative taboo words. Despite the expected reductions in statistical power, the model-fits explained the data in the same way as before. Importantly, in this re-analysis, we obtained two best-fitting models ('Target-only' and 'Probe and Target'), both of which still did not involve any of the association-memory parameters. Hence, differences in valence of the negative words and taboo words cannot explain the observed differences in effects on association-memory across experiments.

8. General discussion

In two experiments, we investigated the influence of arousal on memory for associations in verbal cued recall. Moderately arousing, negative words exerted opposing effects on item- (enhancing) and association- (impairing) memory, and higher-arousing taboo words no longer modulated association-memory at all and only affected item-memory (enhancing). Thus, we found that arousal never enhanced association-memory.

8.1. Convergence with prior studies and the additional contribution of our modeling approach

Interestingly, our results are in accordance with those of both Zimmerman and Kelley (2010) and Guillet and Arndt (2009), studies that came to opposing conclusions about arousal influences on association-memory. Zimmerman and Kelley (2010) also found worse recall of pure negative word pairs compared to pure neutral word pairs (NN < nn), similar to our findings in Experiment 1. Guillet and Arndt (2009), using only mixed pairs, found better cued recall for neutral words that were encoded with taboo words compared to those encoded with moderately arousing, negative or with neutral words (their Experiments 2A–2C), similar to our findings in Experiment 2. Note that Guillet and Arndt (2009) also tested cued recall of neutral words previously paired with moderately arousing, negative (but not taboo) words (similar to our mixed pairs in the Negative group). Contrasting our result of impaired association-memory for neutral words from Nn/nN pairs, they found cued recall was never significantly different from that of neutral words previously paired with other neutral words. We can only speculate about the reasons of this discrepancy, but suggest that floor effects (with average recall rates between 1.4 and 2.6 words in negative–neutral and neutral–neutral conditions in Experiments 2A and 2B, respectively) might have obscured any further impairment of cued recall in the neutral-negative condition. Notably, their Experiment 2C, notwithstanding these floor effects, indeed showed lowered cued recall in the negative–neutral compared to the neutral–neutral condition, similar to our findings in the Negative group, although this difference was not statistically significant.

Based on a superficial examination of Zimmerman and Kelley (2010) and Guillet and Arndt's (2009) results with taboo words, and our own data, one might conclude that arousal levels influence association learning in a non-linear fashion: Moderate arousal being detrimental (NN < nn in Experiment 1; also Zimmerman & Kelley, 2010) and high arousal being beneficial (TT > nn in Experiment 2; also Guillet & Arndt, 2009) for the encoding of associations involving arousing materials. However, based on our modeling results, this account can be ruled out. Moderately arousing, negative words impaired cued recall by reducing association-memory *per se*, whereas higher-arousing taboo words improved cued recall, but only by acting on single words, regardless of what other word they were paired with. Taboo words were more effective cued recall probes and better retrieved as targets; however, they did not improve association-memory *per se*. In other words, although Zimmerman and Kelley (2010) and Guillet and Arndt (2009) came to conflicting conclusions regarding the effect of arousal on association-memory, the results from both of these studies are consistent with our own findings. Importantly, these prior studies included only a subset of the possible word-pairings we included here (only NN and nn in Zimmerman & Kelley, 2010; only Tn, Nn, and nn in Guillet & Arndt, 2009). As such, these studies were missing some of the additional conditions that we found to be diagnostic of association- versus item-level effects of arousal.

8.2. The influence of taboo-specific item properties on memory processes

Our study was not designed to compare potential differences in effects on memory other than arousal between negative and taboo words. Therefore, our results cannot directly 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 sought clarity on this question by conducting a Supplementary multidimensional scaling (MDS) analysis on word norms that were published by Janschewitz (2008) after we had carried out part of the present experiments. The details are reported in the Supplementary materials. Briefly, this database contains subjective ratings on seven dimensions (personal use, familiarity, imageability, arousal, valence, tabooness, and offensiveness) for seven types of words pre-selected to be "taboo" ($n = 92$), "emotionally valenced," subdivided into four sets by high versus low arousal and positive versus negative valence ($n = 46$ each) or "emotionally neutral," subdivided into a related set (all household objects and activities) and an unrelated set ($n = 92$ each). A three-dimensional solution adequately represented the original ratings; however, none of these dimensions clustered the words along a clear arousal dimension. Word types differed significantly from each other in Dimension 1: Taboo words differed significantly from all other word types and both types of negative words differed from all positive and neutral words. There were no 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. Thus, Dimension 1 was not simply related to arousal. If it had been, taboo words would have been followed by high-arousing negative and high-arousing positive words, and these in turn by low-arousing negative/positive words as well as neutral words. We also would have expected significant differences between high- versus low-arousing negative words as well as high- versus low-arousing positive words. In the other two dimensions, taboo words clustered together with negative and positive words against neutral words (Dimension 2) or clustered together with negative and neutral words against positive words (Dimension 3). This suggests that taboo words bear similarity not only to negative, high-arousing words, but other word types as well. We therefore suggest that taboo words are likely perceived as different from other types of words in properties that include, but are not limited to arousal.

Speculating on such differences, we suggest that taboo words are not *just* more arousing than our moderately arousing, negative words from Experiment 1, but that they also represent socially unacceptable or forbidden concepts, or 'tabooness'. According to previous studies in this area, tabooness is defined as a measure of how upsetting the word is to *people in general*. In contrast, offensiveness is defined as the extent to which a participant found the use of the word *personally* offensive (Janschewitz, 2008; Jay, 1992; Jay & Janschewitz, 2008). Additionally, Bertels, Kolinsky, and Morais (2009) suggest that taboo words "are highly arousing but also shocking" and that "shock value is an intrinsic characteristic of [...] taboo words."

Interestingly, taboo-specific effects that cannot be instead be attributed to arousal have been reported in recent attention studies using taboo words. For example, Mathewson, Arnell, and Mansfield (2008) conducted a combined attentional blink–memory study using various types of emotional words. "Threat" words in their study were subjectively rated as similarly arousing to a set of taboo words; however, taboo words still drew significantly more attentional resources than threat words and they were also better recognized. Similarly, Bertels, Kolinsky, and Morais (2010) found significant spatial attention biases towards taboo words compared to other types of words, including negative non-taboo words that were rated as even more arousing than the taboo words in their study. These two results suggest that taboo words can influence attention and memory in ways that cannot be explained simply as effects of arousal alone but may be due to an additional property that could be what subjective judgements of 'tabooness,' 'offensiveness,' or 'shock value' measure.

One possibility is that taboo words may be more distinctive than neutral words, even when matched for word frequency and for frequency of occurrence within the task (as was done here), due to these taboo-specific item properties. Results from the best-fitting models converge with this notion: Taboo words were better probe items than neutral words (model parameter $p > 1$ in Experiment 2; see also Fig. 3a) while negative and neutral words were equivalent as probe items (model parameter $p = 1$ in Experiment 1, Negative group). Note that taboo words were also better target items than neutral words (model parameter $t > 1$ in Experiment 2); however this was true for negative words as well (Experiment 1, Negative group).

8.3. Controlling for semantic cohesiveness

The closer semantic cohesiveness of arousing words can inflate or even subsume item-memory enhancements due to arousal (Talmi & Moscovitch, 2004). However, the results of our Category group in Experiment 1 imply that effects of arousal on association-memory are not just due to semantic cohesiveness. Replicating Talmi and Moscovitch (2004)—but within a cued recall task—we found no difference between final free recall of negative and categorized words (i.e., item-memory, cf. Fig. 1). Furthermore, our modeling results also showed that negative and categorized words equally enhanced target-item retrievability (the t parameter from our model-based approach; cf. Table 3). However, the modeling results further showed between-group differences in the r_1 parameter denoting impairment/enhancement of memory for pure negative (or categorized) pairs relative to mixed pairs: In the Negative group, r_1 was significantly less than 1, suggesting an impairment of association-memory due to arousal while, in the Category group, r_1 was significantly greater than 1, suggesting that memory for associations was enhanced by semantic cohesiveness.

Our subjective ratings of the words indicated no differential cohesiveness amongst negative compared to categorized words; however, we found higher cohesiveness of negative words (*mean* $cos(\theta) \pm sd = .230 \pm .128$) than categorized words (*mean* $cos(\theta) \pm sd = .153 \pm .145$) in our additional LSA-based analyses of the word pools. Even if we had further increased the semantic cohesiveness amongst our category words, to match that of the negative words (according to LSA), for example by using only one single category instead of four different categories, we would expect our modeling results to simply be exaggerated relative to our current findings. That is, with a stronger manipulation of semantic cohesiveness in the category words, we would simply expect r_1 to be even *more* above than 1 than it is currently. Thus, the impairment of association memory found in the Negative group cannot be caused by any potential differences in semantic cohesiveness across the word pools.

8.4. Final free recall as a test of item-retrievability

Free recall is not a pure measure of item-memory. For example, associative cues account for a large amount of the variance in free recall data (e.g., Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981). In our first experiment, we found that NN pairs were recalled worse than nn pairs, even though CC pairs were not impaired relative to nn pairs. In light of our result, we suggest that emotional arousal impairs the learning of specific item–item associations, even though the emotional properties of an item can be used to cue other similarly emotional items (i.e., semantic cohesiveness; see also our discussion of Siddiqui & Unsworth, 2011, in Experiment 1). Additionally, considering that participants in our experiments intentionally studied items as pairs, we would expect an even greater contribution of associative cueing in our final free recall measure than in a typical free recall experiment. Nonetheless, we included our final free recall task, after our cued recall tests, as an additional way to infer effects of item properties on item-retrievability that is less dependent on associative retrieval cues than cued recall. In this respect, our final free recall measure worked as desired and provided evidence of an enhancement of item-memory due to arousal, in spite of the empirical finding of impaired memory for NN pairs relative to nn pairs.

8.5. Lack of temporal asymmetries in the influence of arousal on attention and memory

Our sequential presentation, although intended to allow equal and full attention to both parts of an association and potentially maximize arousal-induced increases in association-memory, may instead have impaired memory for items preceding or following an arousing word, similar to findings in item-memory studies (Bornstein, Liebel, & Scarberry, 1998; Ellis, Detterman, Runcie, McCarver, & Craig, 1971; Hadley & MacKay, 2006; Hurlemann et al., 2005; Johnson et al., 2005; MacKay et al., 2004; Miu et al., 2005; Runcie & O'Bannon, 1977; Schmidt, 2002; Strange et al., 2003). Thus, sequentially presenting each word may have prevented a potential enhancement of association-memory from emerging. However, our data speak against this possibility: If the presentation of a negative or a taboo word had overshadowed attention to and encoding of a subsequent neutral word, we should have observed systematic differences in the recall of the two types of mixed pairs. For example, in Nn pairs, hypothetically, the initial presentation of the negative word would have influenced encoding of the subsequent neutral word. Conversely, in nN pairs, the initial neutral word would have had no influence on subsequent encoding of the negative word. Thus, recall of the neutral words from Nn pairs would have been lower than recall of the neutral words from nN pairs. As can be seen in Fig. 2a (see also Fig. 2c for comparable results with taboo words), this was not the case: Negative and taboo words were better recalled than neutral words in either type of mixed pair with negligible recall asymmetries between the pair types.

9. Conclusion

Although our manipulation of arousal in both experiments increased item-memory, with increases only in the taboo manipulation exceeding those due to semantic cohesiveness, we never found an enhancement of association-memory due to arousal. Our findings replicate and extend previous results of impaired cued recall due to moderately arousing, negative words and enhanced cued recall due to higher-arousing taboo words. Our modeling results uniquely contribute to the extant literature on arousal and association-memory since they could identify the locus of action of such opposing effects: moderately arousing, negative words reduced association memory *per se*, and taboo words increased cued recall only by modulating item-memory. Features of taboo words other than arousal might account for this effect and suggest caution in the use of taboo words to simply manipulate arousal levels. Our results strongly suggest that association-memory is not enhanced by arousal, and that previous results may have been driven by enhanced memory for items, rather than memory for associations.

Acknowledgments

We thank Bridgette Gerson for helping to select stimuli for Experiment 2. We also thankfully acknowledge the help of Pablo Piña Poujol and Jody Arndt for their advice in multidimensional scaling. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and the Alberta Ingenuity Fund.

Appendix A. Experiment 1 word pools

Negative words.

ABUSE	DEATH	HOMICIDE	POVERTY
ACCIDENT	DECEIT	HORROR	PRISON
AFRAID	DEFEAT	HOSTILITY	ROBBERY
ALONE	DEPRESSED	HURRICANE	SAD
ANGER	DESPAIR	HURT	SHRIEK
ARROGANT	DESPISE	INFECTION	SICK
ASSAULT	DISASTER	INSULT	SIN
BLOOD	DISTRESS	JAIL	TERRIBLE
BOMB	EVIL	KNIFE	TERROR
BULLET	FAILURE	LIE	THIEF
BURDEN	FEAR	MALICIOUS	TORTURE
BURIAL	FIGHT	MISERY	TRAGEDY
BURN	FIRE	MURDER	TROUBLE
CRISIS	GRIEF	OFFEND	VICTIM
CRUEL	HATE	PAIN	WEAPON
DANGER	HELL	PANIC	WOUND

Categorized neutral words.

Driving	*School*	*House*	*Business*
BRAKE	COURSE	APARTMENT	BALANCE
DRIVER	DESK	APPLIANCE	BANKER
ENGINE	DISCIPLINE	ATMOSPHERE	BUDGET
MECHANIC	ESSAY	CARPET	CAPITAL
MILEAGE	EVALUATE	CELLAR	CARD
MOTOR	EXAMINE	DECK	CLIENT
PASSENGER	GRADE	DRESSER	CONTRACT
ROAD	LESSON	INHABITANT	CUSTOMER
ROUTE	LIBRARY	MORTGAGE	DIVIDEND
SIGNAL	LOCKER	PORCH	EXECUTIVE
STOP	MARK	PROPERTY	FIRM
TRAILER	PUPIL	RENT	INSURANCE
TRUCK	REGISTER	SHELTER	POUND
TRUNK	REPORT	TERRITORY	RECEIPT
TURN	RULER	WALL	SECRETARY
WHEEL	SCHOLAR	YARD	SELL

Random neutral words.

ALTITUDE	DEEP	JOURNAL	RATTLE
BARREL	DIVISION	LENS	REACTION
BAT	ELBOW	MAGNET	RECENT
BEAT	EQUAL	MELT	SEARCH
BOARD	FARM	MESSENGER	SELECT
BOAT	FLOAT	METAL	SHIP
BOTTLE	FOUNTAIN	METHOD	SOUP
CANDIDATE	FUNCTION	MIMIC	SPRAY
CELL	GRAPH	MIRROR	STATEMENT
CLAY	HANDFUL	MONTH	TACK
CLOUD	HIDDEN	PASSAGE	THUMB
COAT	ICE	PATTERN	TOOL
COLOR	ILLUSION	PENDULUM	TOOTH
COMBINE	INQUIRY	PHASE	TRAIN
CURRENT	IVORY	PLANET	UMBRELLA
DECISION	JACKET	POSTER	WHISTLE

Appendix B. Experiment 2 word pools

Taboo words.

AMPUTATE	DESTROY	LUST	SEX
ANUS	DILDO	NAKED	SHIT
AROUSED	DISGUST	NIPPLE	SHOCK
ASSHOLE	DISMEMBER	NUDE	SLAVE
BARF	EJACULATE	ORGASM	SLUT
BASTARD	ERECTION	ORGY	SPERM
BEATEN	EROTIC	PENETRATE	STAB
BITCH	FART	PENIS	SUFFER
BREAST	FOREPLAY	PISS	SUICIDE
CARESS	FUCK	PUBIC	SURGERY
CLIMAX	GAY	PUSSY	TESTICLE
CLITORIS	GRAVE	RAGE	THREAT
COCK	HERPES	RUBBER	TITS
CONDOM	INCEST	SCREAM	TUMOR
COPULATE	KILL	SEMEN	VAGINA
CUNT	LESBIAN	SENSUAL	WHORE

Neutral words.

ACADEMY	DRILL	JUNGLE	RADIO
AXE	ENABLE	KINGDOM	REPAIR
AXLE	ETERNAL	LOOP	SADDLE
BARREL	EVOKE	LUXURY	SCREW
BELT	EXPERT	MALLET	SLIM
BINDER	FADE	MANAGE	SPONSOR
BLIMP	FAINT	MARBLE	SURF
BOAR	FENCE	MARS	TILT
CABIN	FLAG	NETWORK	TRIBUTE
CHEEK	GARAGE	NIBBLE	TRIM
CHISEL	GRADE	NOTICE	VALID
CLAMP	HAMMER	OMELET	VERBAL
COLUMN	HANGAR	ORANGE	WEDGE
CROWBAR	HATCHET	PERCH	WIRE
DAZZLE	HERD	PLANE	WRENCH
DISCUSS	ICICLE	POND	ZIPPER

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2012.04.001.

References

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General, 134*, 258–281.

Anderson, A. K., Wais, P. E., & Gabrieli, J. D. E. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 1599–1604.

Berntsen, D., & Rubin, D. (2004). Cultural life scripts structure recall from autobiographical memory. *Memory & Cognition, 32*, 427–442.

Bertels, J., Kolinsky, R., & Morais, J. (2009). Norms of emotional valence, arousal, threat value and shock value for 80 spoken French words: Comparison between neutral and emotional tones of voice. *Psychologica Belgica, 49*, 19–40.

Bertels, J., Kolinsky, R., & Morais, J. (2010). Emotional valence of spoken words influences the spatial orienting of attention. *Acta Psychologica, 134*, 264–278.

Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12*, 119–131.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 379–390.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59.

Bradley, M.M., & Lang, P.J. 1999. Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report No. C-1. Gainesville, FL, The Center for Research in Psychophysiology, University of Florida.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.

Buchanan, T. W., Etzel, J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic arousal and semantic relatedness on memory for emotional words. *International Journal of Psychophysiology, 61*, 26–33.
Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language, 59*, 183–199.
Burnham, K. E., & Anderson, D. R. (2002). *Model selection and multimodel inteference* (2nd ed.). New York: Springer-Verlag.
Burnham, K. E., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*, 261–304.
Christianson, S. A. (1992). Emotional-stress and eyewitness memory – A critical-review. *Psychological Bulletin, 112*, 284–309.
Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64*, 119–132.
Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1199–1212.
Davidson, P. S. R., McFarland, C. P., & Glisky, E. L. (2006). Effects of emotion on item and source memory in young and older adults. *Cognitive, Affective, & Behavioral Neuroscience, 6*, 306–322.
Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion*, 5–11.
Ellis, N. R., Detterman, D. K., Runcie, D., McCarver, R. B., & Craig, E. M. (1971). Amnesic effects in short-term memory. *Journal of Experimental Psychology, 89*, 357–361.
Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: A cross-platform experiment-programming library. *Behavior Research Methods, 39*, 950–958.
Guillet, R., & Arndt, J. (2009). Taboo words: The effect of emotion on memory for peripheral information. *Memory & Cognition, 37*, 866–879.
Hadley, C. B., & MacKay, D. G. (2006). Does emotion help or hinder immediate memory? Arousal versus priority-binding mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 79–88.
Hare, R. D. (1973). Orienting and defensive responses to visual stimuli. *Psychophysiology, 10*, 453–464.
Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*, 85–98.
Hurlemann, R., Hawellek, B., Matusch, A., Kolsch, H., Wollersen, H., Madea, B., et al. (2005). Noradrenergic modulation of emotion-induced forgetting and remembering. *Journal of Neuroscience, 25*, 6343–6349.
Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behaviour Research Methods, 40*, 1065–1074.
Jay, T. (1992). *Cursing in America*. Philadelphia: John Benjamins.
Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *American Journal of Psychology, 12*. 183-103.
Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research, 4*, 267–288.
Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., & Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: Prefrontal correlates of refreshing a just-activated representation. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 339–361.
Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition, 30*, 823–840.
Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 701–722.
Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*, 1169–1180.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.
Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476–490.
MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition, 32*, 474–488.
Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language, 63*, 46–63.
Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., & Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience, 18*, 614–625.
Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*, 114–133.
Mathewson, K. J., Arnell, K. M., & Mansfield, C. A. (2008). Capturing and holding attention: The impact of emotional words in rapid serial visual presentation. *Memory & Cognition, 36*, 182–200.
Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language. <http://www.neuro.mcw.edu/mcword/>.
Miu, A. C., Heilman, R. M., Opre, A., & Miclea, M. (2005). Emotion-induced retrograde amnesia and trait anxiety. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1250–1257.
Onoda, K., Okamoto, Y., & Yamawaki, S. (2009). Neural correlates of associative memory: The effects of negative emotion. *Neuroscience Research, 64*, 50–55.
Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinions in Neurobiology, 14*, 198–202.
Philips, L. (1990). Hanging on the metaphone. *Computer Language Magazine, 71*, 238–244.
Philips, L. (2000). The double metaphone search algorithm. *C/C++ User's Journal, 18*, 38–43.
Pierce, B. H., & Kensinger, E. A. (2011). Effects of emotion on associative recognition: Valence and retention interval matter. *Emotion, 11*, 139–144.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition, 14*, 79–94.
Runcie, D., & O'Bannon, R. M. (1977). An independence of induced amnesia and emotional response. *American Journal of Psychology, 90*, 55–61.
Schmidt, S. R. (2002). Outstanding memories: The positive and negative effects of nudes on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 353–361.
Schmidt, S. R., & Saari, B. (2007). The emotional memory effect: Differential processing or item distinctiveness? *Memory & Cognition, 35*, 1905–1916.
Siddiqui, A. P., & Unsworth, N. (2011). Investigating the role of emotion during the search process in free recall. *Memory & Cognition, 39*, 1387–1400.
Siegrist, M. (1995). Effects of taboo words on color-naming performance on a Stroop test. *Perceptual and Motor Skills, 81*, 1119–1122.
Smith, A. P. R., Henson, R. N. A., Rugg, M. D., & Dolan, R. J. (2005). Modulation of retrieval processing reflects accuracy of emotional source memory. *Learning & Memory, 12*, 472–479.
Strange, B. A., Hurlemann, R., & Dolan, R. J. (2003). An emotion-induced retrograde amnesia in humans is amygdala- and β-adrenergic-dependent. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 13626–13631.
Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition, 32*, 742–751.
Touryan, S. R., Marian, D. E., & Shimamura, A. P. (2007). Effect of negative emotional pictures on associative memory for peripheral information. *Memory, 15*, 154–166.
Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavior Research Methods, Instruments, & Computers*, 6–11.
Zimmerman, C. A., & Kelley, C. M. (2010). "I'll remember this!" Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language, 62*, 240–253.

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.

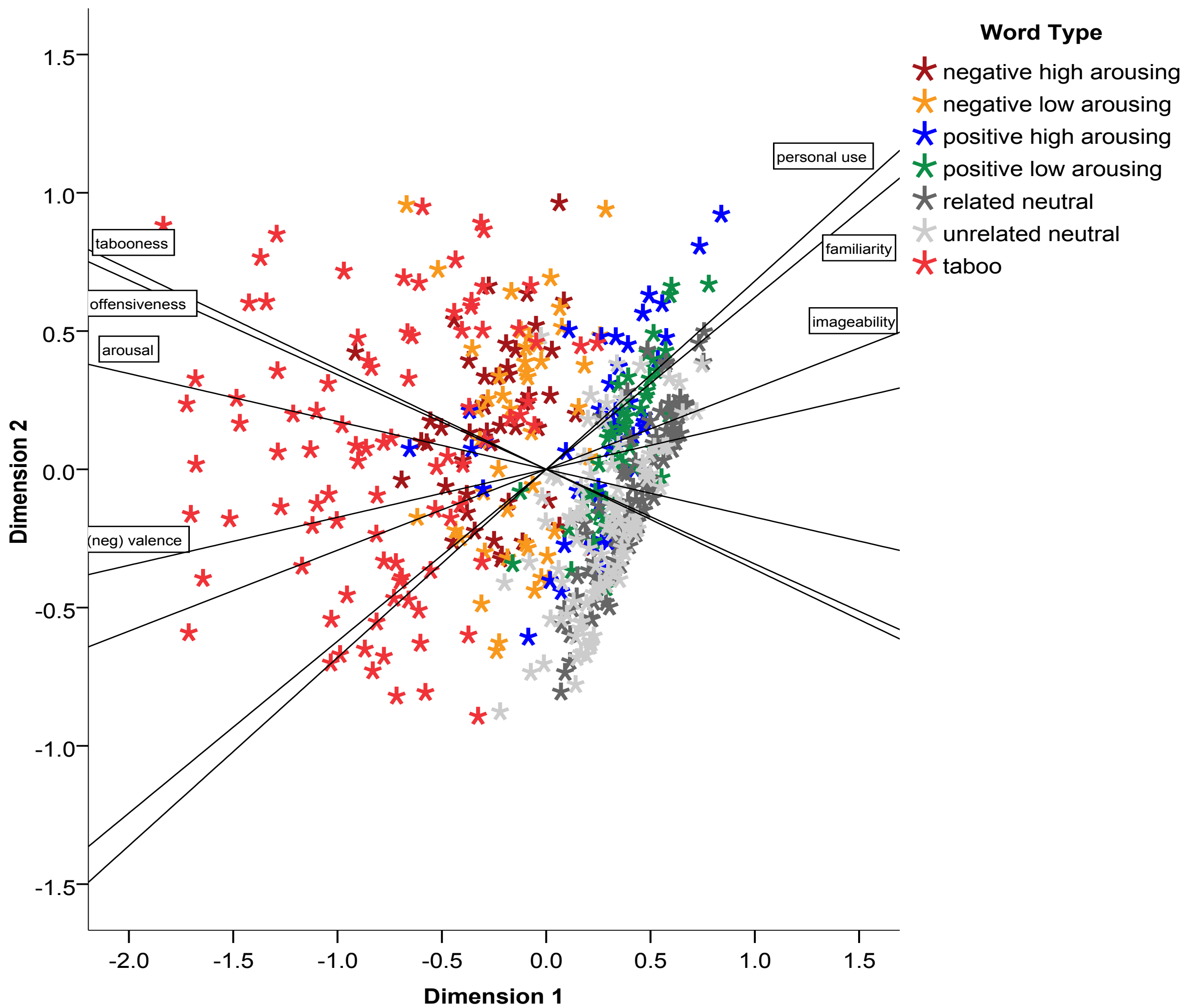

Figure S1: Placements of the 460 words in MDS Dimension 1 and Dimension 2, separated by word type. Averaged beta scores were used to plot regression vectors onto the original word ratings.

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].

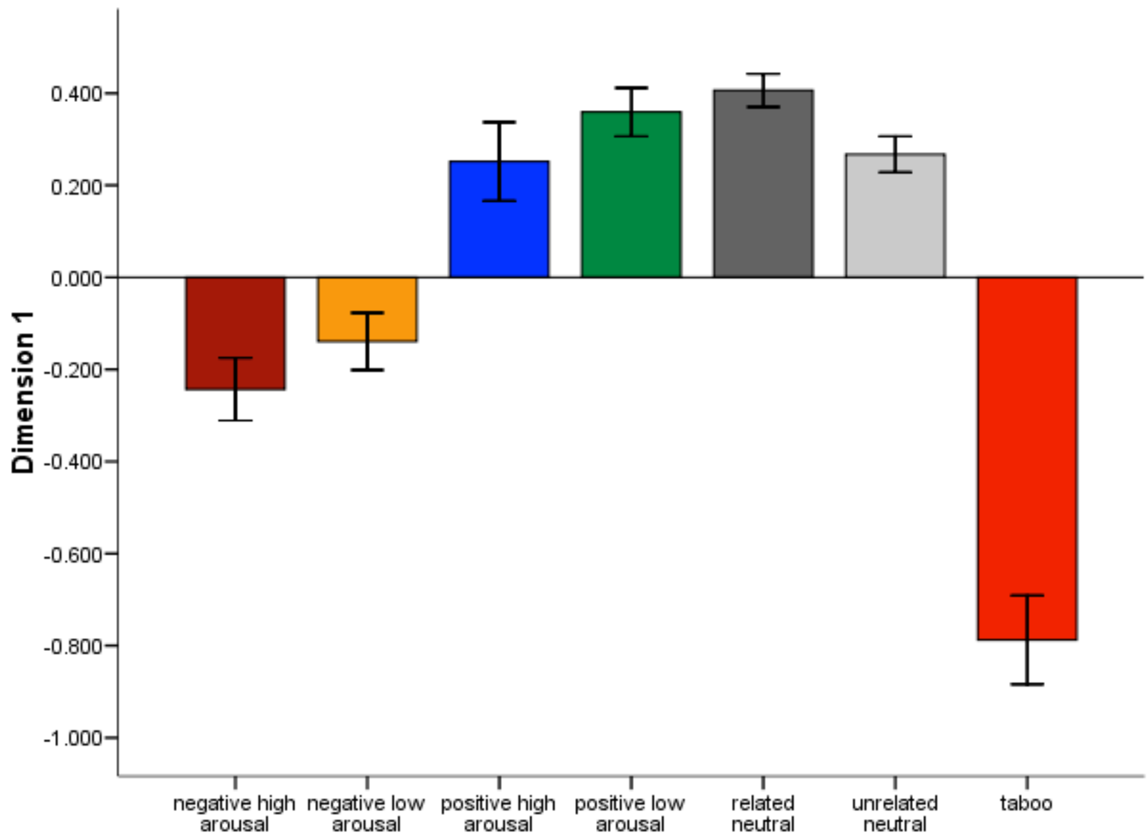

Figure S2: Average common space scores of words in Dimension 1, separated by pre-determined word types. Error bars represent 95% confidence intervals.

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 [$t(128) = 4.28$, $p < 0.05$].

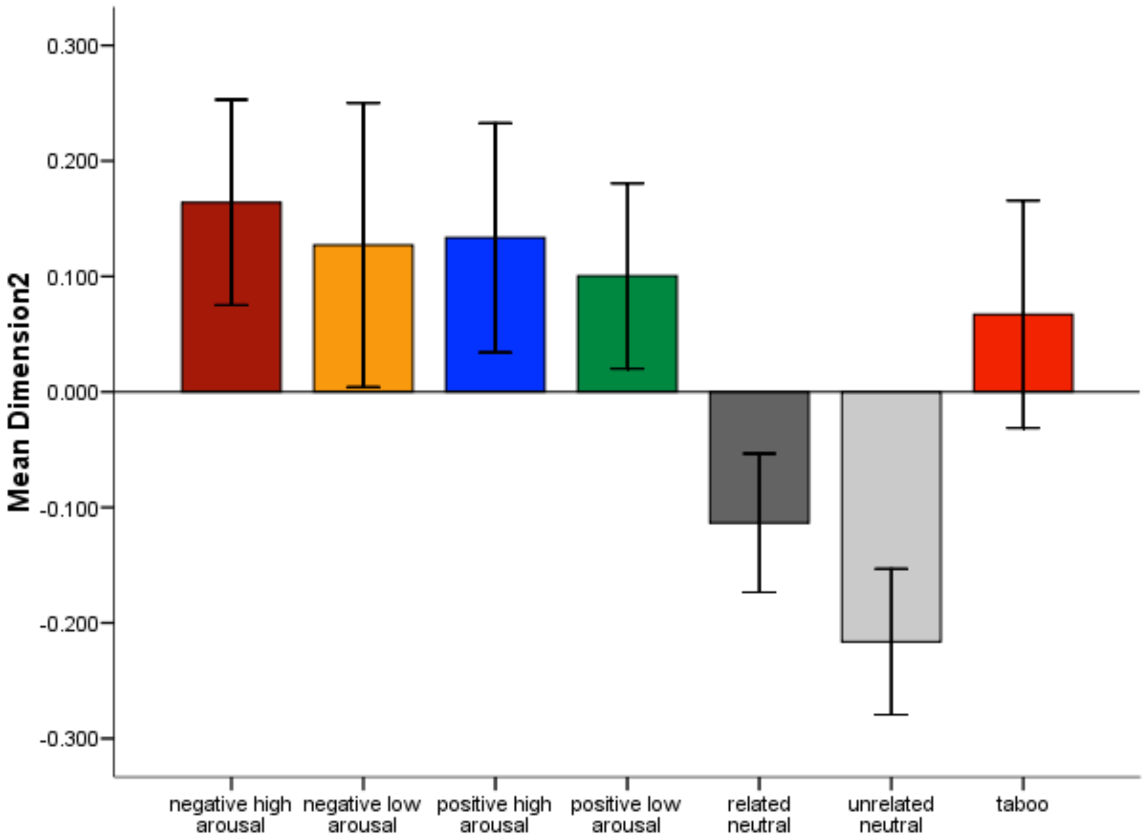

Figure S3. Average common space scores of words in Dimension 2, separated by pre-determined word types. Error bars represent 95% confidence intervals.

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 [$t(453) = 4.82$, $p < 0.01$].

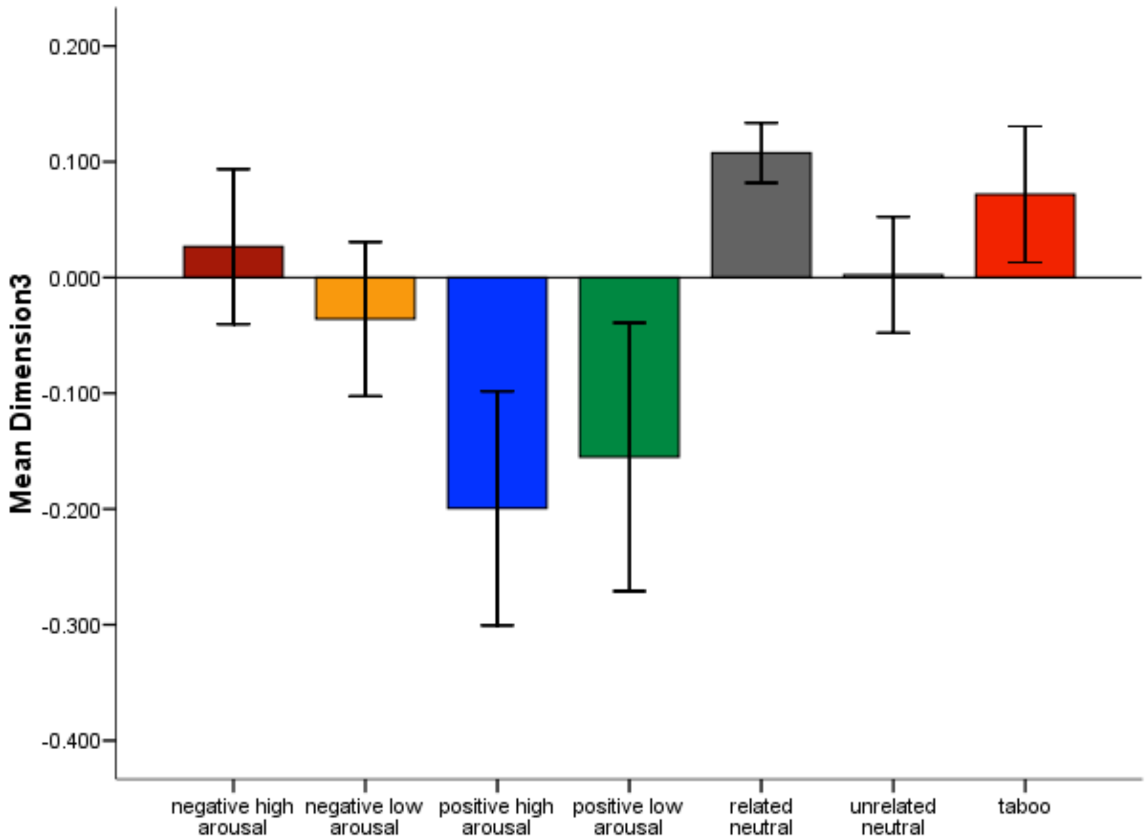

Figure S4. Average common space scores of words in Dimension 3, separated by pre-determined word types. Error bars represent 95% confidence intervals.

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.