THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2017
VOL. 70, NO. 4, 793–810
http://dx.doi.org/10.1080/17470218.2016.1167925

Routledge
Taylor & Francis Group

Shock and awe: Distinct effects of taboo words on lexical decision and free recall

Christopher R. Madan[a,c*], Andrea T. Shafer[b,d*], Michelle Chan[a] and Anthony Singhal[a,b]

[a]Department of Psychology, University of Alberta, Edmonton, AB, Canada; [b]Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada; [c]Department of Psychology, Boston College, Chestnut Hill, MA, USA; [d]Institute of Gerontology, Wayne State University, Detroit, MI, USA

ABSTRACT

Taboo stimuli are highly arousing, but it has been suggested that they also have inherent taboo-specific properties such as tabooness, offensiveness, or shock value. Prior studies have shown that taboo words have slower response times in lexical decision and higher recall probabilities in free recall; however, taboo words often differ from other words on more than just arousal and taboo properties. Here, we replicated both of these findings and conducted detailed item analyses to determine which word properties drive these behavioural effects. We found that lexical-decision performance was best explained by measures of lexical accessibility (e.g., word frequency) and tabooness, rather than arousal, valence, or offensiveness. However, free-recall performance was primarily driven by emotional word properties, and tabooness was the most important emotional word property for model fit. Our results suggest that the processing of taboo words is influenced by distinct sets of factors and by an intrinsic taboo-specific property.

ARTICLE HISTORY
Received 11 March 2015
Accepted 17 February 2016
First Published Online 3 May 2016

KEYWORDS
Emotion; Taboo words; Lexical decision; Free recall; Arousal; Word frequency; Tabooness

Taboo stimuli are defined as "a class of emotionally arousing references with respect to body products, body parts, sexual acts, ethnic or racial insults, profanity, vulgarity, slang, and scatology" (Jay, Caldwell-Harris, & King, 2008). They are thought to be more arousing, more "shocking", and more memorable than other types of emotional information (Bertels, Kolinsky, & Morais, 2009; Buchanan, Etzel, Adolphs, & Tranel, 2006; Janschewitz, 2008; Madan, Caplan, Lau, & Fujiwara, 2012). Furthermore, some studies suggest that taboo stimuli may also possess an inherent taboo-specific property, referred to as "tabooness", which is often defined as how inappropriate (i.e., offensive or shocking) the stimulus is to the general population (Bertels et al., 2009; Janschewitz, 2008; Jay, 1992; Jay & Janschewitz, 2008; Madan et al., 2012). While it has been established that taboo words are processed differently from neutral and non-taboo emotional words, it is currently unclear which stimulus properties (e.g., arousal, valence, tabooness, word frequency, age of acquisition, imageability, word length) contribute to the differences in cognitive processing.

While many would agree on the label of some words as being "*taboo*" (e.g., see Appendix for the list of words used in the current study), there is a dearth of research into the influence of taboo words on cognitive processes. As eloquently described by Jay (2009), "taboo words are sanctioned or restricted on both institutional and individual levels under the assumption that some harm will occur if a taboo word is spoken". Indeed, people are generally hesitant to use taboo words, which also makes their occurrence more striking when they are uttered. (See Allan & Burridge, 2006; Jay, 1992, for detailed discussions of taboo words.) Despite the uniqueness of taboo words as a category, it is unclear how co-varying properties of these words may influence how they are cognitively processed. Generally, taboo words are low in word frequency and high in

*These authors contributed equally to this work.

CONTACT Christopher R Madan madanc@bc.edu Department of Psychology, Boston College, McGuinn 512A, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA

emotional arousal (although they may be positive or negative in valence; Jay, 2009); however, some have suggested that a more critical difference between taboo words and other word types may be due to a taboo-specific word property (e.g., Bertels et al., 2009; Janschewitz, 2008; Jay, 2009; Madan et al., 2012).

Extant evidence showing that taboo words are processed differently from other word types largely comes from lexical decision (LD) and free recall tasks. Regarding LD, evidence suggests that taboo words impair lexical accessibility, as indicated by slower response times (e.g., Geer & Bellard, 1996; Thomas & LaBar, 2005; Williams & Evans, 1980; but see MacKay et al., 2004) when compared to neutral words. Explanations for this impairment are largely based on the emotional properties of taboo words. For example, slower response times for taboo and threatening words/images have been attributed to a perceptual defence mechanism, where participants respond slower to the stimuli due to a strategy of avoiding processing (Geer & Bellard, 1996; McGinnies, 1949; Williams & Evans, 1980), or due to a motor suppression mechanism, as in a temporary freezing response (Azevedo et al., 2005; Estes & Verges, 2008; Fox, Russo, Bowels, & Dutton, 2001; Madan, 2013; Wilkowski & Robinson, 2006). Furthermore, and as found with emotional words (see Fischler & Bradley, 2006, for a review), another possible explanation is that taboo words are more likely to undergo automatic processing (e.g., as indicated by electroencephalography, EEG; Begleiter & Platz, 1969), which, depending on task demands, is related to slower response times (Frings, Englert, Wentura, & Bermeitinger, 2010). However, recent research has shown either no or only small effects of emotional word properties on LD when the non-emotional properties of emotional word stimuli are controlled for (Larsen, Mercer, & Balota, 2006). Therefore, it is possible that non-emotional properties may also account for all or most of the effects of taboo words on LD. Moreover, if taboo stimuli possess an inherent taboo-specific property, then the degree of the contribution of emotional properties to LD performance may differ for taboo compared to non-taboo emotional words.

Previous studies have not directly compared the impact of taboo and non-taboo emotional words on LD. Thomas and LaBar (2005) indirectly made this comparison when they examined the effects of implicit priming on LD response times for extremely high arousing negative (HAN, "taboo"), low arousing negative (LAN), and neutral words. In their paradigm, participants studied HAN/taboo, LAN, and neutral words by deciding whether a word was concrete or abstract. Each word was presented in the centre of the screen for 3000 ms, during which the participant made their abstract/concrete judgment. After the completion of this initial "study phase" task, participants then performed an LD task in which the words that were in the study phase were presented as word stimuli along with new HAN/taboo, LAN, and neutral words. For the new words in the LD task there was a main effect of word type (i.e., emotion) on LD performance. Since the main focus of Thomas and LaBar (2005) was the effect of emotional arousal on implicit priming, they were only interested in LD performance to the repeated word stimuli. Thus, a post hoc assessment of this main effect for the new words was not reported. However, their results showed that in addition to significantly slower abstract/concrete judgements for HAN/taboo words, implicit priming was greater for HAN/taboo words than for neutral words, with LAN words falling in between. Taken together, these findings suggest that compared to neutral words, early semantic processing of HAN/taboo words upon initial exposure is impaired, but after repeated exposure these words receive a beneficial boost in early processing. Therefore, even though these data suggest that processing differences exist between taboo and non-taboo emotional words, open questions remain regarding the outcome of a direct comparison of taboo versus non-taboo emotional words on LD, and the impact of emotional properties (including a taboo-specific property) on LD when non-emotional word properties are controlled for.

Another line of evidence supporting the suggestion that taboo words are processed differently from other word types is that they are remembered better (e.g., Buchanan et al., 2006; Jay et al., 2008; Kensinger & Corkin, 2003; MacKay et al., 2004; Madan et al., 2012). Regarding emotional memory, semantic relatedness has been shown to explain some effects of emotion on free recall probability. Emotional words are more likely to be closely associated with one another than are neutral words (Buchanan et al., 2006; Madan et al., 2012; Talmi & Moscovitch, 2004), and increased semantic relatedness is related to increased retrievability (Tulving & Pearlstone, 1966). In an experiment investigating the effect of emotion on association memory, Madan et al. (2012) compared the similarities and differences between taboo and non-taboo emotional words by performing a multidimensional scaling analysis on words selected from a commonly

used 'word' database with taboo norms (Janschewitz, 2008). Their results suggest that taboo stimuli may influence memory through taboo-specific mechanisms beyond what can be explained by arousal (i.e., tabooness). Thus, it remains unknown whether semantic relatedness is also important in explaining increased free recall for taboo relative to non-taboo emotional words. Although this study suggested a boost in memory for taboo words, it was unable to directly test the effect of a taboo-specific mechanism on item memory. That is, it examined the effect of emotional arousal on memory for word pairs (i.e., association memory) and not memory for singular words (i.e., item memory). Therefore, it remains unclear how specific item properties contribute to this effect.

Although taboo words can differ from other word stimuli on a number of properties (e.g., emotional words are also generally lower in written word frequency, are longer in word length, and have smaller orthographic neighbourhoods; Fackrell, Edmonson-Jones, & Hall, 2013; Larsen et al., 2006), extant research with taboo words has shown less experimental control than studies with more commonly used stimuli. For instance, Jay (2009) discusses that several researchers have suggested that part of the enhancing effect of taboo words on memory may be due to word frequency (e.g., Kensinger & Corkin, 2003; LaBar & Phelps, 1998; Williams & Evans, 1980). However, Jay (2009) argues that taboo words are not actually low in frequency; instead they are rather common in spoken but not written frequency. Fortunately, we now have several ways to measure these effects that were not available to previous researchers: (a) Word frequency estimates can now be obtained from television and film subtitles, which are more representative of language use than estimates obtained from written corpuses (Brysbaert & New, 2009); (b) the Janschewitz (2008) normative word database included measures of familiarity and personal use (see Materials) and were explicitly designed to incorporate taboo words within the database; therefore this database may be more relevant to everyday lexical accessibility than frequency counts obtained from a corpus.

Previous studies investigating the impact of taboo stimuli on cognitive processing have done so taking a categorical (univariate) approach for analysis and, in the case of LD, have only examined differences in processing relative to neutral word stimuli. Therefore, it remains unknown whether taboo words are processed differently from non-taboo emotional words at an early processing stage (e.g., word identification, as tested by speed and accuracy in LD) or whether this difference only emerges during later, more elaborative processes (e.g., memory retrieval, as tested by later free recall). Thus, the first goal of the current investigation was to implement an univariate analysis to (a) replicate previous research by showing that LD for taboo words is slowed, and memory is enhanced when compared to neutral words; and (b) extend previous research by comparing the impact of taboo relative to non-taboo emotional (negative and positive) words on LD.

While a categorical approach is effective in examining overall differences in general emotional categories (e.g., taboo vs. neutral), it does so at the expense of a more in-depth understanding of the role of other word properties, as well as the role of the affective dimensions constituting the emotion categories (i.e., valence, arousal, tabooness). Moreover, given the myriad of word properties that can differ between taboo words and other types of words, it is unclear what factors predominately contribute to observed differences in lexical accessibility and retrievability. Therefore, the second goal of this investigation was to implement a multivariate statistical approach to better evaluate the influence of a variety of word properties on LD and free recall performance. Furthermore, by controlling for non-emotional word properties we sought to determine whether emotional properties provided a unique contribution to LD and free recall.

Based on the extant literature, we made the following three predictions. First, consistent with previous research (Geer & Bellard, 1996; Thomas & LaBar, 2005; Williams & Evans, 1980), we predicted that using a categorical approach to assess the influence of emotional word categories on task performance would show a main effect of word type such that taboo words would delay LD response times and enhance free recall probability relative to neutral words. Second, based on previous findings from non-taboo emotional words (Larsen et al., 2006), we predicted that using a multivariate approach would show that non-emotional word properties best explain LD performance. Third, based on previous findings concerning the impact of emotion on memory (see Dolcos & Denkova, 2008, for review) and the presence of taboo-specific mechanisms (e.g., Janschewitz, 2008; Jay, 1992), we predicted that using a multivariate approach to examine the influence of item properties on free recall performance

would show that free recall is influenced by both emotional and non-emotional word properties.

In Experiment 1, we examine these issues using pre-selected word types (i.e., taboo, negative, positive, neutral) from the Janschewitz (2008) database. In Experiment 2, we used two large pre-existing databases to try to replicate the findings from Experiment 1 regarding the effect of emotional and non-emotional word properties on LD.

EXPERIMENT 1

Experiment 1 consisted of a lexical decision task followed by a free recall test to examine how different emotional word categories influenced LD and free recall performance. Using these data, we sought to replicate and extend previous investigations of the effect of taboo word processing on LD and free recall by (a) comparing taboo, non-taboo emotional, and neutral words in a univariate context, and (b) examining the influence of emotional word properties on LD and free recall after controlling for non-emotional word properties.

Method

Participants

Thirty-nine undergraduate students at the University of Alberta ($M \pm SD = 19.9 \pm 3.0$ years; 27 females; 34 right-handed) participated for partial credit in an introductory psychology course. All participants were required to have learned English before the age of six. Participants gave written informed consent prior to beginning the study, which was approved by the University of Alberta Ethical Review Board.

Materials

Four 40-word lists were used in this experiment: one list of highly arousing, taboo words; one list of moderately arousing, positive valence words; one list of moderately arousing, negative valence words; and one list of emotionally neutral words. All of the words were selected from the Janschewitz (2008) normative word database based on subjective ratings, as well as within-list similarity and orthographic frequencies. See Table 1 for the word property statistics and the Appendix for the specific words used in the experiment.

Subjective ratings from the Janschewitz (2008) normative word database, made on a scale of 1 to 9 (9 being the highest), were used to select words for the experiment. Several ratings were used, including: arousal (how exciting/attention-grabbing the word is), valence (9 = positive; 1 = negative), tabooness (how offensive the word is to people in general), offensiveness (how personally offensive the word is to the rater themselves), familiarity (how often rater has encountered the word), personal use (how often rater personally uses the word), and imageability (conduciveness to mental imagery), as well as the number of letters and syllables. Words were selected such that the words differed in specific word properties (i.e., arousal, valence, tabooness, offensiveness), but were matched on the remaining properties. Briefly, taboo words were more arousing than the other three word types, and positive and negative words were equally more arousing than the neutral words. Positive words were the highest in valence, negative words had the lowest valence, and neutral words were of an intermediate valence. The selected taboo words had a mean valence between that of the negative and neutral words and statistically differed from both. The taboo words were rated highly in tabooness, with the other three words having minimal tabooness. Negative words were slightly (but significantly) higher in tabooness than the positive and neutral words. The offensiveness ratings followed the same pattern as tabooness.

Word frequency (occurrences in the English language, per million words) were obtained from the $SUBLTEX_{US}$ Corpus (Brysbaert & New, 2009), which is based on subtitles from films and television series and contains word frequency counts for 51 million words. This recent database has been shown to explain more variance in lexical decision response times than extant word frequency databases, including Kučera–Francis and CELEX. This database also includes a measure of contextual diversity, which represents how many documents within a corpus a word is found within. Importantly, contextual diversity has been found to account for additional variance in lexical decision response times above variance that could be explained by word frequency (Brysbaert & New, 2009). Only one word was not found in $SUBLTEX_{US}$ ("skank"), for which we substituted in word frequency and contextual diversity values that corresponded to one occurrence within the database, as suggested by Brysbaert and New (2009; also see Brysbaert & Diependaele, 2013). The word lists did not statically differ in word frequency or contextual diversity ($ps > .1$).

Age of acquisition (AoA; age at which a word is learned) has also been shown to be an important predictor of lexical decision response times (Brysbaert & Cortese, 2011; Juhasz, Yap, Dicke, Taylor, & Gullick,

Table 1. Word property statistics for each list.

Word properties		Taboo	Positive	Negative	Neutral	Between-list statistics
Emotional	Arousal	4.38 (0.87)	2.74 (0.55)	2.85 (0.57)	1.60 (0.26)	taboo > positive = negative > neutral
	Valence	4.06 (0.93)	6.37 (0.62)	3.32 (0.46)	5.06 (0.23)	positive > neutral > taboo > negative
	Absolute Valence	1.11 (0.72)	1.37 (0.60)	1.68 (0.46)	0.14 (0.19)	negative > taboo = positive > neutral
	Tabooness	4.59 (1.19)	1.10 (0.14)	1.65 (0.37)	1.07 (0.11)	taboo > negative > positive = neutral
	Offensiveness	2.31 (0.70)	1.03 (0.03)	1.36 (0.19)	1.05 (0.08)	taboo > negative > positive = neutral
Non-emotional	Imageability	5.54 (1.62)	5.06 (2.30)	4.97 (1.68)	5.85 (2.06)	neutral > negative
	Familiarity	5.01 (0.94)	5.06 (0.60)	4.98 (0.94)	4.75 (0.85)	
	Personal use	3.82 (1.10)	4.36 (0.67)	4.11 (0.99)	4.13 (0.95)	positive > taboo
	Letters	5.38 (1.48)	5.58 (1.36)	6.00 (1.47)	5.80 (1.16)	
	Syllables	1.73 (0.72)	1.70 (0.69)	1.78 (0.62)	1.85 (0.48)	
	Word frequency	14.93 (19.15)	14.38 (17.70)	16.50 (24.58)	10.70 (10.66)	
	Contextual diversity	5.41 (6.82)	4.95 (5.07)	5.93 (7.71)	4.06 (3.44)	
	Age of acquisition	8.71 (2.67)	7.97 (2.62)	7.48 (2.45)	7.93 (2.15)	taboo > negative
	Semantic similarity	0.07 (0.14)	0.13 (0.10)	0.10 (0.11)	0.08 (0.09)	

Note: Mean ratings are shown with standard deviations in parentheses. Between-pool statistical differences are listed in the last column, based on at $p < .05$; pools do not differ unless otherwise stated. See text for further details about each measure.

2011). Here we obtained AoA ratings from a recent database developed by Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012), which contains ratings for 30,000 words obtained using Amazon Mechanical Turk. This database contained AoA ratings for all but four of our 160 words (3 taboo, 1 positive). The taboo words had a higher AoA than the negative words, but all other pairwise between-list comparisons were not significant ($ps > .1$).

Within-list semantic similarity was calculated using the latent semantic analysis method (LSA; Landauer & Dumais, 1997). Briefly, this method can be used to model the similarity between two words within semantic space, measured as LSA cos(θ). Semantic space was modelled using the TASA (Touchstone Applied Science Associates Inc.) corpus ("General reading up to 1st year college" semantic space with all 300 available factors). Note that six of the taboo words were not in the TASA corpus and were excluded from this analysis. The four word lists did not differ in mean LSA cos(θ) ($ps > .1$).

A total of 80 pronounceable non-words were generated with the LINGUA non-word generator (Westbury, Hollis, & Shaoul, 2007). Word length was matched to the words (non-words were generated using a Markov chaining length of three; except for three-letter non-words, which had a chaining length of two).

Procedure

The study consisted of three sequential tasks: (a) lexical decision, (b) free recall, and (c) affective ratings. The entire study was completed in 1 hour.

Lexical decision. Participants were presented with letter strings and were asked to judge whether the letter string was a word or non-word. This task took approximately 40 min to complete.

Letter strings were presented for 200 ms in white "Courier New" font, on a black background, in the centre of a computer screen. If the letter string was judged to be a word, participants were instructed to press "P" on the keyboard with their right index finger, and if judged to be a non-word, to press "Q" with their left index finger. Trials were separated with an inter-trial interval ranging from 3 to 5 s. Trials were presented in four blocks of 60 trials each: 10 words of each of the four word types (taboo, positive, negative, neutral) and 20 non-words. The order that the letter strings were presented was randomized for each participant.

Free recall. Immediately following the lexical decision task, participants had 5 min to recall as many words as they could from the experiment. Participants were instructed to type in a word and press the "Enter" key, after which the screen was cleared, and the participant was allowed to type in another word. Spelling errors were corrected using a method adapted from Madan, Glaholt, and Caplan (2010), using the UNIX spellchecking program aspell; repeated responses were only counted once.

Affective ratings. Immediately following the recall task, participants rated all of the words first for arousal, and then for valence. Words were presented one at a time on the computer screen, along with the respective 9-point Self-Assessment Manikin

diagram (SAM; Bradley & Lang, 1994). Participants were instructed to click on the appropriate figure to make their rating. For the valence ratings, the portrait version of the SAM scale was used (Suk, 2006). This portrait version is identical to the standard SAM scale, but enlarges the SAM images to focus just on the face of the manikin. The presentation order of the words was randomized in each rating task.

Data analysis

Participants who responded correctly to less than 85% of the lexical decision trials were excluded from all further analyses ($N = 3$). Only responses made between 200 ms and the individual participant's mean plus 3 standard deviations were included in the analysis (1.76% trials excluded). Incorrect responses were excluded (5.29% trials). Due to machine error, six participants were unable to complete the affective ratings task; thus, rating analyses are based on the responses of the remaining 30 participants (23 females).

For categorical analysis, effects were considered significant based on an alpha level of .05. All analyses of variance (ANOVAs) had one factor, word type (four levels: taboo, positive, negative, neutral), and are reported with Greenhouse–Geisser correction for non-sphericity where appropriate. Post hoc *t* tests were corrected for multiple comparisons using the Holm–Bonferroni method (see Abdi, 2010, for details). Response time analyses were conducted on the within-subject log-transformed mean response time for each word type to allow for parametric statistics (e.g., ANOVAs) on response time.

To obtain a more fine-grained understanding of the influence of affective properties on word processing, we additionally conducted item-wise analysis examining the relationship between each of the word properties obtained from the Janschewitz (2008) normative word database, in combination with the normative databases for word frequency, contextual diversity, and age of acquisition (Brysbaert & New, 2009; Kuperman et al., 2012). In particular, this approach allows us to quantify the relative contributions of arousal, tabooness, and other word properties on lexical decision response times and free recall probabilities. To assess the relationship between each word property and performance, we first performed zero-order correlations between each property, LD response time, and proportion recalled. A more detailed item-wise analysis was then conducted using multiple linear regression to examine the difference between the impact of emotional and non-emotional word properties on the model fit explaining performance. A set of regression models were made for each task. For each set we first aggregated the word properties into two categories: emotional word properties (arousal, valence, absolute valence, tabooness, offensiveness) and non-emotional word properties (imageability, familiarity, personal use, number of letters, number of syllables, word frequency, contextual diversity, age of acquisition). To establish a baseline model that would be used for subsequent model comparison, we then entered all of the word properties for each category into a multiple linear regression model.

For the LD model set, the dependent variable was the log-transformed mean response time. For the free recall model set, the dependent variable was the proportion recalled. Different models were then conducted where appropriate to determine which emotional word properties (if any) of interest (i.e., valence, arousal, tabooness) resulted in a better model fit for explaining task performance. Model fit was assessed via Bayesian information criterion (BIC), which 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, and smaller (i.e., more negative) BIC values correspond to better model fits (Burnham & Anderson, 2002, 2004). We opted to implement this model comparison approach to assess the individual contribution of word properties to the model fit explaining performance as it removes reliance on interpreting beta coefficients that are confounded due to complex suppression effects that arise in stepwise regression when many independent variables are included in the model (Beckstead, 2012; Shieh, 2006).

Results

Confirmatory analyses

To verify that the pre-selected word types corresponded to participants' emotional perceptions of the stimuli, we compared participant's ratings data to the normative ratings from the Janschewitz (2008) database. Word type had a significant effect on both participants' ratings of arousal and valence [arousal: $F(2, 61) = 11.32$, $p < .001$, $\eta_p^2 = .28$; valence: $F(2, 55) = 126.52$, $p < .001$, $\eta_p^2 = .82$]. As expected, participants' arousal and valence ratings correlated with those obtained from the Janschewitz (2008) normative database [arousal: $r(158) = .82$, $p < .001$; valence: $r(158)$

= .95, $p < .001$]. Additionally, these strong correlations support our use of the other ratings from the Janschewitz database (e.g., tabooness, personal use).

Categorical analyses showed that taboo words slowed lexical decision and enhance free recall

We first tested for difference in lexical decision and free recall performance based on the word type (i.e., emotional word category) to see whether we could replicate prior findings. As shown in Figure 1A, word type had a significant effect on lexical decision response time, $F(2, 67) = 8.13$, $p < .001$, $\eta_p^2 = .19$. Specifically, response times for taboo words were greater than response times for positive and neutral words (both *ps* < .05, Cohen's $d > 0.22$), but did not differ relative to negative words after correcting for multiple comparisons ($d = 0.13$). Negative words had greater response times than neutral words ($p = .009$, $d = 0.11$). All other comparisons were not significant, including differences between positive and negative words ($d = 0.10$).

As shown in Figure 1B, word type had a significant effect on recall probability, $F(2, 65) = 147.85$, $p < .001$, $\eta_p^2 = .81$. Recall probability was measured as the proportion of words recalled, relative to words responded to correctly in the lexical decision task.[1] Taboo words were recalled more than any other word type (*ps* < .05, *ds* > 2.8). All other comparisons were not significant.

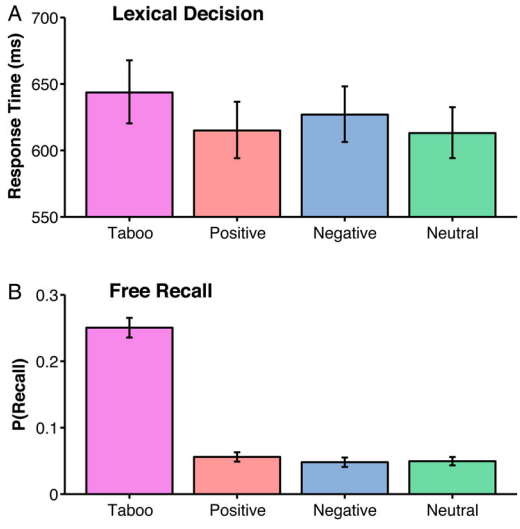

Figure 1. Results from the word type categorical analyses. (A) Mean (±*SEM*) response times in the lexical decision task. (B) Mean (±*SEM*) probability of recall from the free recall task. To view this figure in colour, please visit the online version of this Journal.

Item-wise analyses showed that non-emotional word properties and tabooness best explain delayed lexical decision

To more precisely measure the effects of the word properties on lexical decision performance, we then implemented an item-wise correlation analysis that examined the relationship between each word property and LD response time (Figure 2). Arousal, tabooness, and offensiveness were emotional word properties significantly related to LD response time, such that increases in these properties resulted in slower response times. While the direction of the relationship for valence (specifically, negative-valence words) was the same as that for arousal, tabooness, and offensiveness, its contribution was only marginally significant (see Table 2). Non-emotional word properties related to an item's lexical accessibility (familiarity, personal use, word frequency, contextual diversity) were related to faster response times in LD, whereas non-emotional word properties related to length (number of letters, number of syllables) were related to slower lexical decision response times.

Table 3 summarizes the various linear regression models and, for significant models, their respective measures of model fitness relative to the baseline model (LD01). The emotional model (LD02) was non-significant, and the R^2 of the model was .066, indicating that only a small portion of the variability was explained. For the non-emotional word properties model (LD03), the R^2 of the model was .326 (p < .001). Here, we found that the non-emotional model explained the data significantly better than the baseline model that included all of the properties (ΔBIC relative to LD01 = −16.51), demonstrating that variability in lexical decision response times was best explained by non-emotional word properties.

Next, in models LD04-06 we tested whether the emotional properties (arousal, tabooness, offensiveness), which significantly correlated with performance (shown in Table 2), affected model fit compared to the non-emotional only model (LD03). We assessed model fit for models LD04–LD06 by adding each emotional property one at a time to the non-emotional model, rather than to drop each one from the baseline model, as the non-emotional model was a better fitting model. While all three of these models performed better than the baseline model, only the model that included tabooness (LD05) out-performed the non-emotional model (LD03; ΔBIC from LD03 = −2.26).

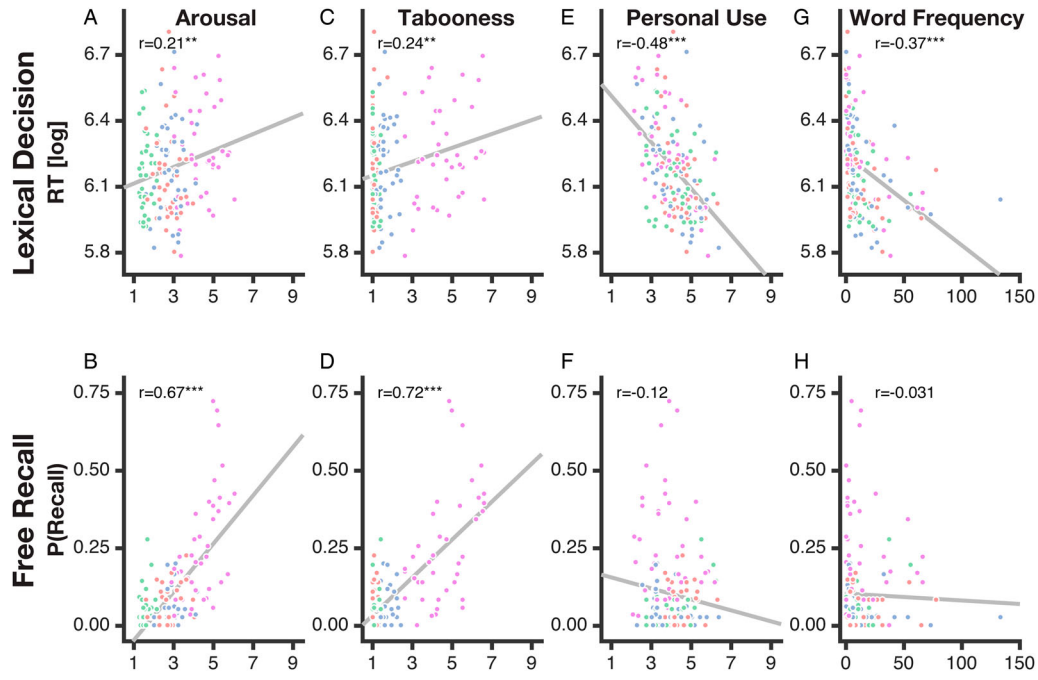

Figure 2. Item-wise correlation analyses of word properties with lexical decision response time (RT) and free recall probability: (A, B) arousal; (C, D) personal use; (E, F) familiarity; (G, H) word frequency. Each dot represents a single word, with the dot colour corresponding to the word type: taboo = purple, positive = red, negative = blue, neutral = green (same as Figure 1). Also see Table 2. $^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. To view this figure in colour, please visit the online version of this Journal.

Item-wise analyses identified a taboo-specific property involved in enhanced free recall

To more precisely measure the influence of the word properties on free recall performance we performed an item-wise correlation analysis that examined the relationship between each word property and free recall probability (Figure 2). All emotional word properties, except valence, significantly contributed to free recall performance.[2] Increases in arousal, tabooness, and offensiveness resulted in enhanced memory performance. Imageability was the only non-emotional word property shown to contribute to enhanced memory performance.

Mirroring the multiple linear regression analysis conducted for lexical decision, we conducted a set of multiple linear regression models using sets of

Table 2. Correlations of word properties with lexical decision response time and free recall probability.

		Experiment 1		Experiment 2	
Word properties		LD RT [log]	FR P(Recall)	ELP LD RT [log]	BLP LD RT [log]
Emotional	Arousal	.21**	.67***	.14***	.07
	Valence	−.14†	−.13	−.08†	−.10*
	Absolute valence	.06	−.04	−.02	.00
	Tabooness	.24**	.72***	.25***	.20***
	Offensiveness	.20*	.57***	.12**	.18***
Non-emotional	Imageability	−.15†	.24**	−.14***	−.13**
	Familiarity	−.41***	.06	−.50***	−.56***
	Personal use	−.48***	−.12	−.53***	−.58***
	Letters	.19*	−.03	.46***	.28***
	Syllables	.16*	.06	.44***	.18***
	Word frequency	−.37***	−.03	−.33***	−.31***
	Contextual diversity	−.42***	−.07	−.40***	−.41***
	Age of acquisition	−.02	.02	.55***	.49***

Note: LD = lexical decision; FR = free recall; RT = response time; ELP = English Lexicon Project; BLP = British Lexicon Project.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 3. Multiple linear regression analyses for Experiment 1.

			Word properties														
			Emotional					Non-emotional									
			J2008										BN2009		K2012		
Behavior		Model	Arousal	Valence	Absolute valence	Tabooness	Offensiveness	Imageability	Familiarity	Personal use	Letters	Syllables	Word frequency	Contextual diversity	Age of acquisition	R^2	ΔBIC vs. all
Lexical decision RT [log]	LD01	All	×	×	×	×	×	×	×	×	×	×	×	×	×	.363***	–
	LD02	Emotional	×	×	×	×	×									.066†	21.75
	LD03	Non-emotional						×	×	×	×	×	×	×	×	.326***	−16.51
	LD04	Non-emotional + arousal	×					×	×	×	×	×	×	×	×	.349***	−16.82
	LD05	Non-emotional + tabooness				×		×	×	×	×	×	×	×	×	**.357***	**−18.77**
	LD06	Non-emotional + offensiveness					×	×	×	×	×	×	×	×	×	.351***	−17.34
Free recall P(Recall)	FR01	All	×	×	×	×	×	×	×	×	×	×	×	×	×	.621***	–
	FR02	Emotional	×	×	×	×	×									.589***	−29.10
	FR03	Non-emotional						×	×	×	×	×	×	×	×	.239***	83.55
	FR04	All − arousal		×	×	×	×	×	×	×	×	×	×	×	×	.610***	−0.52
	FR05	All − tabooness	×	×	×		×	×	×	×	×	×	×	×	×	**.602***	**2.75**
	FR06	All − offensiveness	×	×	×	×		×	×	×	×	×	×	×	×	.621***	−5.03

Note: RT = response time; ×s denote word properties that were included in the respective regression model. References for the databases where the word property measures were obtained are abbreviated: J2008 denotes Janschewitz (2008); BN2009 denotes Brysbaert and New (2009); K2012 denotes Kuperman et al. (2012). Bolded rows highlight the models that specifically tested the influence of tabooness on lexical decision and free recall, respectively.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

word properties to explain free recall probability (Table 3). As before, we initially started with three models: all of the word properties (FR01), only emotional word properties (FR02), and only non-emotional word properties (FR03). The model that contained all of the word properties performed remarkably well, accounting for 62.1% of the total variance ($p < .001$). The emotional model was able to account for nearly all of this variance with only five word properties and was a better fit than the baseline model (FR02: $R^2 = .589$, $p < .001$, ΔBIC relative to FR01 = −29.10), whereas the non-emotional model fit was poorer than the baseline model (FR03: $R^2 = .239$, $p < .001$, ΔBIC relative to FR01 = 83.55).

From Table 2, the emotional word properties that had significant correlations with free recall probability were tabooness, arousal, and offensiveness. As the emotional model was better fitting than either the baseline or non-emotional models, we took a slightly different approach from that used above for the LD data to determine which emotional word property had the largest contribution to explaining memory performance when non-emotional word properties were also considered. To test this, we constructed three additional models, where each contained all of the word properties except arousal, tabooness, or offensiveness, respectively (FR04–FR06 in Table 3). In other words, after accounting for all of the other word properties we could, we asked whether the additional inclusion of tabooness would result in a better fitting model. Removing arousal had no effect on the model fit (FR05: $R^2 = .610$, $p < .001$, ΔBIC relative to FR01 = −0.52). The model fit improved after removing offensiveness (FR05: $R^2 = .621$, $p < .001$, ΔBIC relative to FR01 = −5.03). In contrast, when tabooness was removed from the set of word properties, the model fit was significantly worse (FR05: $R^2 = .602$, $p < .001$, ΔBIC relative to FR01 = 2.75). These results suggest that of these three word properties, tabooness is important in producing the best model to explain memory performance. Thus, it is apparent that arousal is not sufficient in explaining how taboo words are processed, and a taboo-specific property (i.e., tabooness) better explains recall behaviour.

Discussion

Experiment 1 identified three main findings. First, we replicated and extended previous work by showing that taboo words resulted in slower LD performance than did neutral and positive words, but their degree of impairment did not differ from that of negative words. Second, tabooness was the only emotional property to uniquely contribute to LD once non-emotional word properties were controlled for (specifically, those related to lexical accessibility). Third, we found that both emotional and non-emotional word properties contributed to free recall performance. In particular, we found that a taboo-specific property operates independently of other emotional properties to best explain recall.

Categorical analyses showed that taboo words impair lexical decision and enhance free recall

Our result showing slowed LD and enhanced free recall for taboo compared to neutral words is consistent with earlier reports comparing the impact of taboo words on early word processing (e.g., Geer & Bellard, 1996; Thomas & LaBar, 2005; Williams & Evans, 1980; but see MacKay et al., 2004). Extending previous work on taboo words and LD, we showed that LD performance for taboo words was slower than that for positive words, but did not differ from that for negative words. This result was not due to differences in arousal as it was equated across the non-taboo emotional word conditions.

As with LD, taboo words had the largest impact on free recall such that the proportion recalled was greater than that for any other word category (e.g., Buchanan et al., 2006; Jay et al., 2008; Kensinger & Corkin, 2003; MacKay et al., 2004). To rule out the possibility that increased memory for taboo words was due to their increased response time in the LD task, we evaluated whether these two measures were related. Performing a correlation between lexical decision response time and free recall probability yielded no relationship, $r(158) = .07$, $p > .1$. This suggests that these two phenomena are distinct and are likely to be driven by unique differences in word properties. However, we explored this idea further to determine whether differences exist in the relationship between immediate and delayed processes as a function of the overarching categories of word properties (i.e., emotional vs. non-emotional). We performed two partial correlations between LD and free recall performance. The first partial correlation, $r(158) = -.155$, $p = .05$, examined the immediate and delayed relationship due to non-emotional properties (i.e., all emotion-related variance was removed). The second partial correlation, $r(158) = .032$, $p = .69$, examined the immediate–delayed relationship due to emotional properties (i.e., all variance related to the non-

emotional word properties was removed). This analysis shows that, at the behavioural level, the influence of emotion on LD and free recall is distinct, whereas the influence of non-emotional word properties is related across tasks such that faster lexical accessibility is linked to increased memory.

Numerous studies examining incidental emotional memory have found a relationship between behaviour during encoding and subsequent memory such that task-irrelevant emotion results in impaired performance on an encoding task. This in turn results in enhanced performance on a subsequent memory task (Kensinger & Corkin, 2003; Pottage & Schaefer, 2012; Shafer & Dolcos, 2012; Talmi, Schimmack, Paterson, & Moscovitch, 2007). The impairment in the encoding task performance is thought to be due to a reallocation of resources away from the task-relevant stimulus features and towards the task-irrelevant emotional aspects of the stimulus. Although we did not find evidence of this immediate impaired–delayed enhanced relationship in an item-wise analysis, it is reasonable to assume that the boost in memory for taboo words should be observable in the encoding data. There are three possibilities as to why we did not observe this relationship. First, individual differences that are collapsed in an item-wise analysis may have washed out the effect. Second, the response time measure may not be sensitive enough to detect an immediate–delayed relationship. Third, previous studies examining this relationship have only used perceptual-based discrimination tasks. Perhaps this behavioural relationship only exists for lower-level perceptual tasks. This idea is indirectly supported by the negative relationship observed between LD response time and free recall for the non-emotional word properties, where faster LD resulted in enhanced memory. Future research using EEG or functional magnetic resonance imaging to examine the relationship between LD and free recall performance for taboo words might be very useful to elucidate the encoding mechanisms responsible for the boost in memory.

Surprising to us and inconsistent with a large body of extant research on emotional memory, we found no differences in free recall between the (non-taboo) emotional word categories and neutral words. There are three reasons why this may have occurred. First, some studies have reported that increases in memory for non-taboo emotional stimuli may be the result of increased semantic similarity. This was controlled for by ensuring that all emotional word categories did not differ in semantic similarity (see Table 2). Therefore increased retrievability of emotional words due to increased relatedness was eliminated. This finding is consistent with that of previous research that has examined memory for taboo words versus conditions where semantic relatedness was manipulated (Jay et al., 2008). Second, the arousal elicited by the taboo words may have interfered with the processing of words presented immediately before and after them (MacKay et al., 2004). Therefore, one possible interpretation of this finding is that processing taboo words interfered with encoding mechanisms of temporally adjacent words, although the length of the temporal window where words preceding or following a taboo word are impacted has not been thoroughly investigated. This explanation is unlikely in the context of the current task parameters as the inter-trial interval ranged from 3 to 5 s. Third, the short delay between the encoding and retrieval tasks may have reduced the likelihood of finding differences between non-taboo emotional and neutral words as the memory-enhancing effect of emotion is maximized with a delay period of at least 20 min (Kleinsmith & Kaplan, 1963). Consequently, another possible (but not mutually exclusive) interpretation of this finding is that the degree of arousal evoked by the taboo words was enough to observe an immediate impact on memory, whereas the ability to observe more subtle effects of arousal on memory may require a longer interval between encoding and retrieval.

Another issue to consider is the type of memory test administered. For instance, Buchanan et al. (2006)—using a longer inter-trial interval (~3 s)—did not observe a significant enhancement of memory for negative versus neutral words, at either an immediate or a 1-hour delayed free recall test. Additionally, Jay et al. (2008) used a self-paced encoding task that presumably resulted in shorter inter-trial intervals than in the current study, and in Buchanan et al. (2006), and also had a 10-min delay between the encoding task and free recall test. Once semantic relatedness was controlled for in the neutral condition, memory for emotional words was not significantly different. Kensinger and Corkin (2003) implemented a recognition paradigm with a stimulus duration of 2 s and no inter-trial interval during the encoding task. There was a 15-min delay between encoding and retrieval, and memory was assessed in separate experiments for source information and for recollection versus familiarity. They found better source

memory for negative than for neutral words, but after controlling for semantic relatedness of the neutral words, increased memory for negative words occurred only for items that were considered to be recalled (based on the remember vs. know response paradigm; Tulving, 1985, 1993). Thus, the main difference between these previous studies in showing a memory enhancement for non-taboo emotional relative to neutral words in the context of taboo words is the use of free recall versus recognition retrieval paradigms. This suggests that increased encoding occurred for non-taboo emotional words, but retrieval processing was impaired due to a taboo-focused orientation dominating the search process during retrieval. Once the search process is re-oriented via a cue away from the taboo information and towards the non-taboo emotional information, enhanced memory for the non-taboo emotional information is preserved. Future research targeting these potential explanations for the findings regarding non-taboo emotional memory in the context of testing memory for taboo stimuli will be necessary for a better understanding of the underlying mechanisms resulting in increased memory for taboo stimuli as well as alterations in non-taboo emotional memory.

Non-emotional word properties and tabooness best explain delayed lexical decision

A univariate analysis reviewed above highlights differences in LD performance due to the emotional category of the word (taboo, positive, negative, neutral). However, previous research looking at differences in LD between emotional and neutral words found that differences due to emotion were eliminated once the variance explained by non-emotional word properties was controlled for (Larsen et al., 2006). Somewhat discrepant with this finding, we found an unique contribution of tabooness after non-emotional word properties were controlled for. This discrepancy is further discussed in Experiment 2.

A taboo-specific property is necessary to best explain memory performance

Overall, our findings are consistent with earlier work looking at memory for taboo words; however, we extend previous findings by showing that after non-emotional word properties are controlled for, the inclusion of a taboo-specific property (as measured here by tabooness ratings) best explained memory performance. The current results indicate that even though taboo words differ from other words on many word properties, tabooness is the primary emotional factor that leads to increased retrievability.

EXPERIMENT 2

While Experiment 1 provided evidence that a taboo-specific word property—that is, tabooness—is important in explaining the effects of taboo words on lexical accessibility and retrievability, it can be argued that the experimental design was too constrained. Specifically, by only presenting 160 words in total, which were selected specifically to maximize between-list differences, the effect of tabooness may have been exaggerated. To address this, we separately analysed LD data from two large, pre-existing databases—the English Lexicon Project (ELP; Balota et al., 2007) and the British Lexicon Project (BLP; Keuleers et al., 2012)—along with all of the word properties we used in Experiment 1, to determine whether we would observe the same LD response time effects in these datasets.

Method

Materials

All 460 words were selected from the Janschewitz (2008) database, along with all of the word properties provided in the database (as described in Experiment 1). For all of the words, we additionally obtained word frequency and contextual diversity measures from Brysbaert and New (2009) and age-of-acquisition ratings from Kuperman et al. (2012). In contrast to Experiment 1, if words were missing from a database, they were excluded from all analyses. The Brysbaert and New (2009) database was only missing 5 words from the Janschewitz database. The Kuperman et al. (2012) database was missing an additional 9 words. Combined, these 14 words from the Janschewitz database that were not found in the Brysbaert and New (2009) and Kuperman et al. (2012) databases were dropped from all analyses.

When available, we obtained the mean LD response times for the 460 words from both the ELP (Balota et al., 2007) and the BLP (Keuleers et al., 2012). The ELP was missing response times for 36 words (all taboo); the BLP was missing response times for 79 words (30 taboo, as defined in Janschewitz, 2008). By design, the BLP only includes mono- and bi-syllable words; of the 79 missing words, 58 were words that were three or more syllables in length (based on the number of syllables measure

included in the Janschewitz, 2008, database). In the final ELP dataset, after additionally removing the 14 words that were missing at least one word property, 418/460 words remained (42 total excluded; 37 taboo, as defined in Janschewitz, 2008). In the final BLP dataset, 376/460 words remained (84 total excluded; 31 taboo excluded). Unsurprisingly, a large proportion of the excluded words were taboo words; however, as the Janschewitz (2008) database contained 92 taboo words, many still remained. Mean response times were log-transformed prior to any analyses being conducted.

Data analysis

Replicating the item-wise analysis in Experiment 1, we examined the relationship between each of the word properties obtained from the Janschewitz (2008) normative word database, in combination with the normative databases for word frequency, contextual diversity, and age of acquisition (Brysbaert & New, 2009; Kuperman et al., 2012). We first performed zero-order correlations between each property and LD response time (see Table 2). We then conducted multiple linear regression to examine the difference between the impact of emotional and non-emotional word properties on the model fit explaining performance. A set of regression models were made with the word properties aggregated into two categories: emotional word properties (arousal, valence, absolute valence, tabooness, offensiveness) and non-emotional word properties (imageability, familiarity, personal use, number of letters, number of syllables, word frequency, contextual diversity, age of acquisition). A baseline model was then established for subsequent model comparison by entering all of the word properties for each category into a multiple linear regression model. Different models were then conducted where appropriate to determine which (if any) emotional word properties of interest resulted in a better model fit for explaining task performance.

Results

Confirmatory analyses

To validate the LD response times that we found in Experiment 1, we conducted correlations between the LD response times that we observed and those obtained from the ELP and BLP databases. Reassuringly, the correlations were relatively strong [ELP: $r(148) = .45$, $p < .001$; BLP: $r(136) = .47$, $p < .001$]. As a comparison, we additionally calculated the correlation between the ELP and BLP, for the subset of words used in Experiment 1, $r(131) = .57$, $p < .001$.

Table 2 reports the zero-order correlations between LD log-transformed mean response time from the ELP and BLP databases with the word properties considered here. Briefly, the results are largely consistent with our findings in Experiment 1, though most of the correlations are significant here due to the large number of degrees of freedom. Note, however, that we did observe additional strong relationships not found in Experiment 1—for example, letters, syllables and AoA positively correlated with response time, and imageability negatively correlated with response time. As Experiment 1 used a more tightly control subset of words, these effects were attenuated in that sample.

Replication of Experiment 1: Item-wise analyses showed that non-emotional word properties and tabooness best explain delayed lexical decision

As in Experiment 1, we calculated sets of multiple linear regression models to determine the contribution of different word properties to model fit assessing LD performance. As shown in Table 4, the full models that included all word properties explained approximately half of the variance in LD response times (ELP01: $R^2 = .489$, $p < .001$; BLP01: $R^2 = .440$, $p < .001$). After constraining the LD datasets based on those in the three word property databases used here (Brysbaert & New, 2009; Janschewitz, 2008; Kuperman et al., 2012), each of these datasets contained more than twice as many words as those used in Experiment 1 [ELP: 2.61; BLP: 2.35] and are inherently based on responses from many more participants. Thus, it is re-assuring that these full models explain even more variance than we were able to in Experiment 1 (Table 3), where mean LD response times are probably noisier (LD01: $R^2 = .363$, $p < .001$).

Consistent with the findings in Experiment 1, considering the emotional and non-emotional models, in both LD datasets, the non-emotional models better explained the data [(ELP02: $R^2 = .106$, ΔBIC relative to ELP01 = 185.58) vs. (ELP03: $R^2 = .461$, ΔBIC relative to ELP01 = −8.22); (BLP02: $R^2 = .07$, ΔBIC relative to BLP01 = 143.46) vs. (BLP03: $R^2 = .419$, ΔBIC relative to BLP01 = −15.57)]. Identical to the approach used in Experiment 1, we assessed model fit by adding each emotional property one at a time to the non-emotional model as the non-emotional model was a better fitting model. For both datasets, adding the tabooness word property to the non-emotional

Table 4. Multiple linear regression analyses for Experiment 2, where lexical decision response times were obtained from the English and British Lexicon Projects.

			Word properties															
			Emotional					Non-emotional										
			J2008										BN2009		K2012			
Behavior		Model	Arousal	Valence	Absolute valence	Tabooness	Offensiveness	Imageability	Familiarity	Personal use	Letters	Syllables	Word frequency	Contextual diversity	Age of acquisition	R^2	ΔBIC vs. all
ELP	ELP01	All	×	×	×	×	×	×	×	×	×	×	×	×	×	.489***	–
Lexical decision	ELP02	Emotional	×	×	×	×	×									.106***	185.58
RT [log]	ELP03	Non-emotional						×	×	×	×	×	×	×	×	.461***	−8.22
	ELP04	Non-emotional + arousal	×					×	×	×	×	×	×	×	×	.469***	−8.35
	ELP05	Non-emotional + tabooness				×		×	×	×	×	×	×	×	×	**.486*** **	**−21.72**
	ELP06	Non-emotional + offensiveness					×	×	×	×	×	×	×	×	×	.472***	−10.18
BLP	BLP01	All	×	×	×	×	×	×	×	×	×	×	×	×	×	.440***	–
Lexical decision	BLP02	Emotional	×	×	×	×	×									.070***	143.46
RT [log]	BLP03	Non-emotional						×	×	×	×	×	×	×	×	.419***	−15.57
	BLP04	Non-emotional + arousal	×					×	×	×	×	×	×	×	×	.422***	−11.33
	BLP05	Non-emotional + tabooness				×		×	×	×	×	×	×	×	×	**.434*** **	**−19.09**
	BLP06	Non-emotional + offensiveness					×	×	×	×	×	×	×	×	×	.428***	−15.39

Note: RT = response time; ELP = English Lexicon Project; BLP = British Lexicon Project; ×s denote word properties that were included in the respective regression model. References for the databases where the word property measures were obtained are abbreviated: J2008 denotes Janschewitz (2008); BN2009 denotes Brysbaert and New (2009); K2012 denotes Kuperman et al. (2012).
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

model was the only emotional word property to improve model fit (ELP05: ΔBIC relative to ELP03 = −13.5; BLP05: ΔBIC relative to BLP03 = −3.52). Importantly, these findings—non-emotional word properties explaining more variance than the emotional word properties, and the model including non-emotional word properties plus tabooness performing the best of all models considered—mirror the findings of Experiment 1 (see LD03 and LD05 in Table 3).

Discussion

Experiments 1 and 2 had consistent findings regarding which word properties best explain LD performance by showing that non-emotional word properties explained the most variance. Furthermore, only the addition of the emotional property tabooness resulted in improved model fit.

Non-emotional word properties and tabooness best explain delayed lexical decision

Our findings are inconsistent with previous research showing no effect of emotional word properties on LD after non-emotional word properties are controlled for. Previous research looking at differences in LD between emotional and neutral words found that differences due to emotion were eliminated once the variance explained by non-emotional word properties was controlled for (Larsen et al., 2006). This discrepancy is probably due to a combination of methodological and statistical differences. Methodologically, Larsen et al. (2006) did not obtain arousal, valence, and tabooness ratings, but rather adopted emotional categories that were previously determined. Statistically, an analysis of covariance (ANCOVA) was used with the non-emotional word properties, word length, frequency, and orthographic neighbourhood as covariates. Thus, it could be argued that our approach is stronger as we controlled for more non-emotional word properties, included more co-varying emotional properties (i.e., tabooness, offensiveness, absolute valence), and implemented a multivariate approach using continuous variables for all of the emotional and non-emotional word properties.

Contrary to our findings, a number of recent reports have shown valence effects in LD after controlling for non-emotional word properties (Estes & Adelman, 2008; Kousta, Vinson, & Vigliocco, 2009; Larsen, Mercer, Balota, & Strube, 2008; Vinson, Ponari, & Vigliocco, 2014). These reports suggest that the influence of valence is categorical and not continuous (Estes & Adelman, 2008; Kousta et al., 2009; Scott, O'Donnel, & Sereno, 2014; Vinson et al., 2014). However, there is disagreement about how emotion words fit into categories with some studies showing the polarity of valence matters (i.e., negative and positive categories), whereas others show that polarity does not matter, and negative and positive words form one group that differ from neutral words (i.e., emotion and neutral categories). Moreover, discord also exists about the directionality of emotion effects, with some showing facilitation effects of both negative and positive words on LD (Kousta et al., 2009; Scott et al., 2014; Vinson et al., 2014), while others show an impairing effect of negative words and either facilitation or no effect of positive words (Estes & Adelman, 2008; Hofmann, Kuchinke, Tamm, Võ, & Jacobs, 2009; Larsen et al., 2006). Although this issue was not part of the main goal of the current study, we explored whether the relationship between emotional word properties and LD performance differed as a function of emotional category (taboo, positive, negative, and neutral). Once non-emotional word properties were controlled for there were no significant partial correlations between any of the emotional word properties and LD response time. This indicates that the impact of category does not make a difference on LD once non-emotional words properties are considered. Furthermore, adopting the approach implemented here in Experiments 1 and 2, we examined differences in model fit between the non-emotional model relative to the baseline model and the non-emotional model plus valences or absolute valence relative to the baseline model. For Experiments 1 and 2, valence measures had poorer model fit than the non-emotional model. As a reminder, the inclusion of arousal or offensiveness had no impact on model fit compared to the non-emotional model, whereas tabooness improved fit. Although we did not find valence effects at the category level, it is possible that valence interacts with another word property to affect performance (Larsen et al., 2008; Scott et al., 2014). Future research comparing regression or structural equation models with different interaction terms between the word properties will be necessary to fully understand the role of valence.

Although we show that valence does not influence a behavioural assessment of early lexical processing, it is possible that differences in neural processing exist during early lexical processing for different categories

of emotional stimuli (Bayer & Schacht, 2014; Kaltwasser, Ries, Sommer, Knight, & Willems, 2013; Schacht & Sommer, 2009). Brain imaging methodology may offer a more sensitive measure of these differences that are not otherwise observable (as they are no longer maintained at the response level). Furthermore, understanding how individual differences (from either emotional or cognitive domains) contribute to the impact of emotion on LD is important to consider as differences due to the emotional properties may be washed when not considering these (Escobar et al., 2013; Sass et al., 2010).

It should be highlighted that our findings are consistent across three independent datasets. Moreover, this consistency is preserved in the presence of large methodological differences. That is, the current study placed an information processing constraint on participants such that each letter string was displayed for 200 ms, whereas studies included in the ELP and BLP databases displayed the letter strings for longer or until a response was made. Here we used two taboo-related word properties to characterize our stimuli: tabooness and offensiveness, where tabooness is a rating of how offensive the word is to people in general, while offensiveness represents how personally offensive the word is to the rater themselves. While it may seem surprising that tabooness, rather than offensiveness, was more strongly related to the LD and free recall effects reported here, this is probably because we used normative ratings for both measures (drawn from Janschewitz, 2008), rather than ratings from the same participants as those that performed the LD and free recall tasks. Taken together, the LD results from the univariate and item-wise analytical approaches show that words with higher tabooness are processed differently from those with low tabooness and that this difference is largely, but not entirely, attributed to non-emotional word properties related to an item's accessibility (familiarity, personal use, word frequency).

CONCLUSION

The current investigation examined whether there are differences in the impact between taboo and non-taboo emotional words on LD and free recall performance and assessed how different word properties (both emotional and non-emotional) influenced LD and free recall performance. As expected, taboo words were associated with slower response times in lexical decision (i.e., lexical accessibility) and higher recall probabilities in free recall (i.e., retrievability); however, different sets of word properties best explained these effects: Lexical decision performance was best explained by non-emotional word properties linked to lexical accessibility (word frequency, familiarity, and personal use); after these were accounted for, only tabooness improved model fit above the non-emotional model. However, memory performance was explained differently. Free recall was best explained by emotional word properties, and of the emotional properties considered, the inclusion of tabooness was necessary to best explain memory performance. Taken together, our results emphasize the importance of considering the variance explained by non-emotional properties when examining the impact of emotion on cognitive processes using word stimuli. Furthermore, they indicate that the processing of taboo words is influenced by distinct sets of factors, and that taboo words are not merely high-arousal emotional words, but also possess an intrinsic tabooness property.

Notes

1. The statistical significance of all free recall analyses remains the same when instead all words presented are used, rather than just those that the participant responded to correctly in lexical decision.
2. Valence did not significantly correlate with free recall even when we calculated the correlation separately for positive- and negative-valence items.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was partly funded by a Discovery grant from the Natural Science and Engineering Research Council (NSERC) of Canada held by A.S.; Canada Graduate Scholarships (Doctoral level) held by C.R.M. and M.C.; and a Dissertation Fellowship held by A.T.S. from the University of Alberta.

ORCID

Christopher R. Madan http://orcid.org/0000-0003-3228-6501

References

Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N. J. Salkind, D. M. Dougherty, & B. Frey (Eds.), *Encyclopedia of research design* (pp. 573–577). Thousand Oaks, CA: Sage.

Allan, K., & Burridge, K. (2006). Forbidden *words*: Taboo and the *censoring of language*. Cambridge, UK: Cambridge University Press.

Azevedo, T. M., Volchan, E., Imbiriba, L. A., Rodrigues, E. ., Oliveira, J. M., Oliveira, L.F., … Vargas, C. D. (2005). A freezing-like posture to pictures of mutilation. *Psychophysiology*, *42*, 255–260.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … , & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.

Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces–a cross-domain comparison. *Frontiers in Psychology*, *5*, 299.

Beckstead, J. W. (2012). Isolating and examining sources of suppression and multicollinearity in multiple linear regression. *Multivariate Behavioral Research*, *47*, 224–246.

Begleiter, H., & Platz, A. (1969). Cortical evoked potentials to semantic stimuli. *Psychophysiology*, *6*, 91–100.

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.

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

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, *64*, 545–559.

Brysbaert, M., & Diependaele, K. (2013). Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behavior Research Methods*, *45*, 422–430.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

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.

Burnham, K. E., & Anderson, D. R. (2002). *Model selection and multimodel inference* (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.

Dolcos, F., & Denkova, E. (2008). Neural correlates of encoding emotional memories: A review of functional neuroimaging evidence. *Cell Science Reviews*, *5*, 78–122.

Escobar, M. J., Rivera-Rei, A., Decety, J., Huepe, D., Cardona, J. F., Canales-Johnson, A., … & Ibañez, A. (2013). Attachment patterns trigger differential neural signature of emotional processing in adolescents. *PLOS ONE*, *8*, e70247.

Estes, Z. & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, Balota (2006). *Emotion*, *8*, 441–444.

Estes, Z., & Verges, M. (2008). Freeze or flee? Negative stimuli elicit selective responding. *Cognition*, *108*, 557–565.

Fackrell, K., Edmonson-Jones, M., & Hall, D. A. (2013). A controlled approach to the emotional dilution of the Stroop effect. *PLOS ONE*, *8*, e80141.

Fischler, I., & Bradley, M. M. (2006). Event-related potential studies of language and emotion: Words, phrases, and task effects. *Progress in Brain Research*, *156*, 185–203.

Fox, E., Russo, R., Bowels, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*, 681–700.

Frings, C., Englert, J., Wentura, D., & Bermeitinger, C. (2010). Decomposing the emotional Stroop effect. *Quarterly Journal of Experimental Psychology*, *63*, 42–49.

Geer, J. H., & Bellard, H. S. (1996). Sexual content induces delays in unprimed lexical decisions: Gender and context effects. *Archives of Sexual Behavior*, *25*, 379–395.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L.-H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, *9*, 389–397.

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

Jay, T. (1992). *Cursing in America*. Philadelphia: John-Benjamins.

Jay, T. (2009). The utility and ubiquity of taboo words. *Perspectives on Psychological Science*, *4*, 153–161.

Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *American Journal of Psychology*, *121*, 83–103.

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research*, *4*, 267–288.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, *64*, 1683–1691.

Kaltwasser, L., Ries, S., Sommer, W., Knight, R. T., & Willems, R. M. (2013). Independence of valence and reward in emotional word processing: Electrophysiological evidence. *Frontiers in Psychology*, *4*, 168.

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.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44, 287–304.

Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, *65*, 190–193.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*, 473–481.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, *9*, 490–493.

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.

Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, *6*, 62–72.

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: importance of word arousal. *Emotion*, *8*, 445–452.

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. (2013). Toward a common theory for learning from reward, affect, and motivation: The SIMON framework. *Frontiers in Systems Neuroscience*, *7*, 59.

Madan, C. R., Caplan, J. B., Lau, C. S. M., & Fujiwara, E. (2012). Emotional arousal does not enhance association-memory. *Journal of Memory and Language*, *66*, 695–716.

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.

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychological Review*, *56*, 244–251.

Pottage, C. L., & Schaefer, A. (2012). Visual attention and emotional memory: recall of aversive pictures is partially mediated by concurrent task performance. *Emotion*, *12*, 33–38.

Sass, S. M., Heller, W., Stewart, J. L., Silton, R. L., Edgar, J. C., Fisher, J. E., & Miller, G. A. (2010). Time course of attentional bias in anxiety: Emotion and gender specificity. *Psychophysiology*, *47*, 247–259.

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition*, *69*, 538–550.

Scott, G. G., O'Donnel, P. J., & Sereno, S. C. (2014). Emotion words and categories: Evidence from lexical decision. *Cognitive Processing*, *15*, 209–215.

Shafer, A. T., & Dolcos, F. (2012). Neural correlates of opposing effects of emotional distraction on perception and episodic memory: An event-related fMRI investigation. *Frontiers in Integrative Neuroscience*, *6*, 70.

Shieh, G. (2006). Suppression situations in multiple linear regression. *Educational and Psychological Measurement*, *66*, 435–447.

Suk, H.-J. (2006). Color and emotion: A study on the affective judgment across media and in relation to visual stimuli. University of Mannheim: PhD dissertation.

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*, 742–751.

Talmi, D., Schimmack, U., Paterson, T., & Moscovitch, M. (2007). The role of attention and relatedness in emotionally enhanced memory. *Emotion*, *7*, 89–102.

Thomas, L. A., & LaBar, K. S. (2005). Emotional arousal enhances word repetition priming. *Cognition and Emotion*, *19*, 1027–1047.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12.

Tulving, E. (1993). Varieties of consciousness and levels of awareness in memory. In A. Baddeley and L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control. A tribute to Donald broadbent* (pp. 283–299). Oxford, UK: Clarendon Press.

Tulving, E. & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381–391.

Vinson, D. P., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition and Emotion*, *28*, 737–746.

Westbury, C., Hollis, G., & Shaoul, C. (2007). LINGUA: The language-independent neighbourhood generator of the University of Alberta. *The Mental Lexicon*, *2*, 271–284.

Wilkowski, B. M., & Robinson, M. D. (2006). Stopping dead in one's tracks: Motor inhibition following incidental evaluation. *Journal of Experiment Social Psychology*, *42*, 479–490.

Williams, L. J., & Evans, J. R. (1980). Evidence for perceptual defense using a lexical decision task. *Perceptual and Motor Skills*, *50*, 195–198.

APPENDIX

Words used in the experiment

Taboo	Positive	Negative	Neutral
anus	admired	assassin	ankle
bastard	angel	blister	banner
blowjob	awed	brutal	barrel
boner	bath	carcass	butter
boobs	brave	chaos	cannon
booty	breeze	cockroach	contents
breasts	bride	depressed	context
butt	bunny	dirt	cord
climax	carefree	divorce	corridor
clit	caress	evil	errand
cock	casino	gloom	fabric
crap	champ	gossip	finger
cum	cozy	headache	gender
dick	elated	horror	glacier
dildo	gift	hurricane	hairpin
fornicate	grateful	illness	icebox
hooker	grin	insane	ink
horny	hopeful	lonely	item
hump	hug	massacre	journal
idiot	joyful	measles	jug
jerk	kitten	mold	lantern
knockers	liberty	morbid	limber
moron	lively	pest	locker
nipples	luxury	pollute	metal
orgasm	queen	poverty	museum
penis	rainbow	quarrel	obey
piss	riches	riot	passage
prick	sailboat	roach	patent
pussy	scholar	rusty	pencil
screw	secure	scared	phase
scrotum	silk	scream	privacy
semen	snuggle	slap	quart
skank	soothe	snake	rattle
snatch	sunrise	stench	spray
suck	sunset	tomb	statue
testicles	thrill	toxic	stiff
tits	tune	trash	teacher
turd	wink	trauma	tower
vagina	wise	urine	trunk
vibrator	wit	waste	writer