Title: Emotion Words in Lexical Decision and Reading Aloud

Authors:

Catherine Mason[a], MRes, catherine.mason@mq.edu.au

Solène Hameau[ab], PhD, solene.hameau@uclouvain.be

Lyndsey Nickels[a], PhD, lyndsey.nickels@mq.edu.au

Affiliations:

[a] School of Psychological Sciences, Macquarie University, Sydney, Australia
[b] Faculty of Psychology and Educational Sciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Corresponding Author:

Catherine Mason

School of Psychological Sciences, Macquarie University, NSW, Australia

catherine.mason@mq.edu.au

Data Availability:

Data, code, preregistration documents and supplementary materials can be accessed at:

https://osf.io/tyjse/files/osfstorage

Number of Tables: 5

Number of Figures: 0

Abstract

Emotion can have facilitative effects on word processing with words that label emotions and/or have strong emotional associations (high valence and/or arousal) showing an advantage on word processing tasks. However, there has been little research on the production of emotion-label words. In this study, we examined emotion effects in production (reading aloud) and recognition (lexical decision), comparing processing of a large set of emotion-label words and non-emotion abstract words. Response latencies to emotion-label and non-emotion words did not differ in lexical decision but emotion-label words showed significantly faster response times in reading aloud. Valence and arousal did not significantly influence response times in either task. We suggest that the presence of an emotion-label advantage only in reading aloud and in the absence of valence/arousal effects points to a fragility of emotion effects. These findings contrast with previous studies which indicate a key role of valence and arousal in word processing.

Keywords

Lexical processing, abstract words, emotion words, word production, valence

Introduction

Research regarding abstract word processing has often highlighted words that describe emotions, such as 'happiness', 'hatred', or 'anxiety', as potentially unique (e.g., Altarriba & Bauer, 2004; Catricalà et al., 2014; Pulvermüller, 2018) with Altarriba and Bauer (2004) even suggesting that emotion words may form a separate subgroup, distinct from other abstract words. This suggestion was driven by their finding of distinct behavioural outcomes for emotion words (i.e., words that specifically label emotions such as 'happiness' and 'anxiety') compared with other abstract words, with emotion words having faster response times on a primed lexical decision task and higher accuracy in free recall. Similar findings of distinct outcomes for emotion words have been found on other tasks. For example, in silent reading, eye-tracking data suggests faster processing of emotion words than neutral words (Knickerbocker et al., 2015); in the Stroop task emotion words but not neutral words show Stroop interference (Sutton & Altarriba, 2008); and in a Flanker task, emotion words are responded to slower than words that do not describe emotions (Zhang et al., 2019).

Further evidence of differences in emotion word processing compared with other abstract words has been reported in clinical studies with evidence of both relative preservation and specific impairment of words that label emotions compared with other word categories. Catricalà et al. (2014) presented evidence for selective preservation of emotion words in people with Alzheimer's Disease (AD) on word association and sentence completion tasks, while 'social relations' and 'other abstract' word categories were impaired (for a similar pattern, see Martin & Fedio, 1983). Better performance on emotion words has also been reported in people with stroke aphasia, for example, showing faster response times in visual lexical decision (Newton et al., 2020) and higher accuracy in reading aloud (Landis et al., 1982). However, these studies have taken a broad classification of emotion words referring to all words with emotional connotations, and not only words that label emotion states. In contrast, Hsieh et al. (2012) reported evidence that emotion words can be selectively impaired in people with behavioural-variant frontotemporal dementia (bvFTD) on a word association task while non-emotion abstract words and concrete words were spared. However, there is the possibility that any deficits in emotion word processing reported in these clinical studies may in fact be associated with emotion processing impairments more generally. For example, Souter et al. (2021) showed that people with semantic variant primary progressive aphasia (sv-PPA) and (stroke) aphasia can show deficits in their ability to identify and categorise emotional facial expressions. As such, further research is needed to distinguish whether deficits in emotion word processing may be the result of impairments of emotion perception and vice versa.

Possible explanations for distinct processing of emotion words compared with other abstract words have been proposed based on specific characteristics that distinguish them from other abstract words (e.g., Altarriba & Bauer, 2004; Altarriba et al., 1999). Specifically, emotion words have been found to have features which are more typically associated with concrete words, such as relatively high imageability (Altarriba et al., 1999; Paivio et al., 1968) and high context availability (Altarriba et al., 1999). There is extensive evidence that imageability facilitates lexical processing (e.g., Balota et al., 2004; Paivio et al., 1968; Strain & Herdman, 1999), and Paivio (1991) explained this effect using dual coding theory according to which all words have a verbal representation, but words with high imageability have an additional non-verbal representation. For high imageability words, these two representations have an additive effect, which is argued to explain their faster processing compared with low imageability words. Similarly, higher ratings of context availability have also been associated with a facilitatory effect on lexical processing (e.g., Colombo & Burani, 2002; Schwanenflugel et al., 1988; Schwanenflugel et al., 1992; Schwanenflugel & Stowe, 1989). Context availability refers to how much a word is tied to a specific context and the extent to which its meaning is reliant on the contexts in which it is used (Schwanenflugel et al., 1988). For example, the word 'happiness' is rated as high in context availability as it is relatively easy to think of a particular context in which it may be used and its meaning is not particularly variable based on context. In contrast, a word such as 'quality' is rated as low in context availability: it is more difficult to think of a context in which it may be used and its meaning changes from context to context (e.g., *'The vase was of high quality'* vs. *'The student's best quality was their positive attitude')*. The high ratings of both imageability and context availability for emotion words distinguish them from other abstract words and these factors are hypothesised to have a facilitatory effect on their processing, similar to that found for concrete words (Altarriba & Bauer, 2004).

An alternative explanation that may account for differences between processing of emotion words and other abstract words, relates to their inherent emotionality. There has been growing interest in how semantic features relating to emotion (i.e., affective features) may impact on lexical processing (see Hinojosa et al., 2020 for a review) and the role these features play in the learning and processing of abstract concepts specifically (e.g., Kousta et al., 2009; 2011; Vigliocco et al., 2009; Vigliocco et al., 2014). To date, there has been a focus on two affective features: arousal (the degree to which a word/concept is associated with activity/passivity), and valence (the degree to which a word/concept is associated with positivity/negativity). In a large-scale study of visual lexical decision, Kousta et al. (2011) showed that words with extreme (very negative or very positive) valence had an advantage over neutrally valenced words while arousal had a marginal effect. In addition, although abstract words had faster response times compared with concrete words when imageability and context availability were controlled, this abstract word advantage was not present when valence and arousal were included in the statistical model. Kousta et al. (2011) explained this effect of affective features within an embodied language account of lexical processing, suggesting that while visual imagery provides sensorimotor grounding for concrete concepts, for abstract words grounding comes from emotional features and associations.

An implication of this account, however, is that a processing advantage would be expected for not only emotion words but also for any words with strong valence and arousal and indeed there is evidence that there is a processing advantage for all words with strong emotional associations, even those that do not refer to emotions themselves (e.g., Kousta et al., 2009, 2011; Vinson et al., 2014). This leads to a key issue facing this research, related to how emotion words are defined. There are two possibilities: either they can be words that label emotions, such as 'happiness', or words that have emotional qualities or associations but are not actually emotions, such as 'violence'. There is some evidence for a difference in the outcomes of behavioural experiments between words that specifically label emotions (emotion-label words) and words with emotional associations (emotion-laden words) (e.g., Simon task, Altarriba & Basnight-Brown, 2011; masked/unmasked primed lexical decision, Kazanas & Altarriba, 2015; eye-tracking in silent reading, Knickerbocker & Altarriba, 2013), as well as in in patterns of neural activity (e.g., Wang et al., 2019; Wu et al., 2021; Wu & Zhang, 2019; Zhang et al., 2019). However, when controlling for more factors (e.g., age of acquisition, imageability, concreteness, mean bigram frequency), Vinson et al. (2014) did not find that emotion-label words differed in lexical decision response times from other abstract words with emotional associations.

It is also important to note that while numerous studies have demonstrated effects of valence and/or arousal, there are some inconsistencies in the specific patterns observed. For example, while some studies have demonstrated a facilitative effect of extremity of valence similar to Kousta et al. (2009; 2011) (e.g., Crossfield & Damian, 2021; Lund et al., 2019; Vigliocco et al., 2014; Yap & Seow, 2014), evidence from other authors suggests that the effect of valence may be directional, with negative words showing a processing disadvantage compared with positive words (e.g., Estes & Aldeman, 2008; Kuperman et al., 2014, Larson et al., 2008; Sutton & Lutz, 2019). Outcomes for arousal have also been variable: some studies have failed to find an effect (e.g., Vinson et al., 2014), others have shown low arousal words to have faster response times compared with high arousal words (Kuperman et al., 2014), while the reverse, high arousal words having faster response times than low arousal words, has also been reported (Kever et al., 2019).

In sum, the current evidence suggests that the presence of an emotional element as part of word meaning can influence lexical processing. However, whether words that specifically label emotions can be considered a distinct category of abstract words with different processing characteristics appears inconclusive. Moreover, the extent to which semantic dimensions including valence, arousal, imageability and context availability act on abstract word processing warrants further exploration. Consequently, in this study, we aimed to address this issue further. We not only extend previous research on lexical decision, but also examine reading aloud to explore the production of emotion words.

As discussed previously, emotion words are reported to be rated higher in imageability than other abstract words and so it is of relevance here that imageability has been shown to have a greater effect in lexical decision than reading aloud (Balota et al., 2004;

Cortese & Khanna, 2007; Ferrand et al., 2011). Meanwhile, to our knowledge, there has been no study looking at the comparative effects of word type (emotion vs. non-emotion) between lexical decision and reading aloud tasks. Indeed, there has been relatively little research on the effects of affective variables (valence/arousal) in production tasks, with the notable exceptions of Estes and Adelman (2008) and Larsen et al (2008), however these studies did not focus on abstract words specifically.

We also address some key limitations of much of the previous research: i) a tendency to use relatively small word sets (due to emotion words being a very restricted word group; e.g., Altarriba and Bauer, 2004 $N =$ 30 per set; Kazanas & Altarriba, 2015, $N =$ 14 per set), reducing power; ii) have limited control of other variables, introducing potential confounds, and/or iii) take a broad definition of emotion words (e.g., Kousta et al., 2011; Vinson et al., 2014). Therefore, we maintained a strict definition of emotion words and used a comparatively large set of 104 items (52 emotion, 52 non-emotion abstract) matched across a large range of variables and statistically controlled for the effects of any unmatched features.

This research therefore aimed to investigate whether adult native English speakers show a difference between emotion words and other abstract words in lexical decision and reading aloud. In addition, we explored the effects on these tasks of those variables that have been hypothesised to influence emotion and abstract word processing (imageability, context availability, extremity of valence, and arousal). In particular, we examined the extent to which any difference between the word sets (emotion vs. non-emotion abstract) may be accounted for by the effects of these variables. If emotion words are distinct in their lexical processing, this has implications for stimulus selection in research using abstract words and may also draw into question whether abstract words can be viewed as a cohesive group. An additional motivation for our study was to further explore the Kousta et al.'s (2011) position that affective features (extremity of valence and arousal) may play a particularly important role in the processing of abstract words and to therefore determine whether this finding could be replicated in lexical decision and whether a similar finding may be evident in reading aloud.

Preregistration documents, raw data, code and supplementary materials can be accessed at https://osf.io/tyjse/files/osfstorage. Supplementary Materials include participant, stimuli and analysis details (S1), norming instructions (S2) and all analysis outputs (S3). Aspects of the analysis plan which were not clearly conveyed in the preregistration document are outlined in S1.

Experiment 1: Lexical Decision

The first experiment addressed two research questions. First (Research Question 1.1), we asked: Are there differences in the response times to emotion and non-emotion abstract words on a lexical decision task? In answering this question, we aimed to extend the previous research (Altarriba and Bauer, 2004; Kazanas & Altarriba, 2015) by using larger sets of stimuli and carefully matching emotion and non-emotion words for variables that have sometimes been uncontrolled. Consistent with previous studies, we predicted that emotion words would have shorter response latencies compared to non-emotion abstract words.

Our second research question (Research Question 1.2) addressed the nature of any processing differences between emotion and non-emotion abstract words by asking: Can any effect of word type (emotion/non-emotion abstract) on lexical decision response times be explained by variables previously shown to influence emotion word processing? Previous research has shown that lexical decision response times are influenced by imageability (e.g., Balota et al., 2004; Ferrand et al., 2011), context availability (Schwanenflugel et al., 1988), valence (e.g., Kousta et al., 2009; 2011; Kuperman et al., 2014; Vinson et al., 2014), and arousal (e.g., Kever et al., 2019; Kuperman et al., 2014). As outlined earlier, it has been suggested that one, or a combination, of these variables may underlie differences in processing emotion and other non-emotion abstract words. If this is the case, any effect of word category (emotion vs. non-emotion abstract) should no longer be apparent when these factors are added to the statistical model. This analysis was to be implemented even if no effect of word type was found to further investigate the effects of these variables on abstract word processing, and in particular, whether there was support for Kousta et al.'s (2009; 2011) position that abstract word processing is strongly influenced by affective variables (extremity of valence/arousal).

Methods

Power Analysis

The number of participants was determined in accordance with Brysbaert's (2019) suggestion that the minimum number required for a two-level study is 52. However, as we would be drawing participants from a student group which includes a large proportion of bilingual speakers and previous studies suggest a possible effect of bilingualism on emotion word use (e.g., Pavlenko, 2008), we aimed to recruit 112 participants in total with a minimum of 56 monolingual speakers to allow for loss of data sets and exclusion of bilingual speakers should (pre-planned) analysis indicate an effect of bilingualism.

Participants

One hundred and twelve native English speakers were recruited from the Macquarie University undergraduate psychology student pool over a single semester and received course credit for their participation. Selection criteria were that participants had no history of communication, language or reading difficulties and had no hearing, eyesight or physical difficulties that would impact on their ability to complete the tasks. Participants provided written consent prior to participating in the experiments and were informed that they may stop participating at any time. This study was conducted with ethics approval from the human research ethics committee at Macquarie University.

Addition of bilingualism to a model with all experimental variables (Analysis 1.2; Model 4 in Analysis 1.2, See Table 2 for model structures) revealed a significant interaction between bilingualism and arousal (Log transformed data: $\beta = 0.01$, SE = 2.54e-03, $t(9937) = -2.33$, $p = .020$; Untransformed data: $\beta = 3.76$, SE = 1.85, $t(9936) = 2.03$, $p = .043$). We therefore excluded bilingual participant data from the analyses, resulting in the inclusion of 67 (monolingual) participants. These participants ranged from 18-53 years of age, with a mean of 20.5 years (standard deviation, SD = 5.40). Forty-four participants identified as female, 19 as male and three as other.

Materials

Word stimuli consisted of 52 emotion and 52 non-emotion abstract words, as well as 52 concrete words which acted as fillers and were not included in the analyses. Emotion words were selected from the emotion categories (including synonyms or near-synonyms) described by Cowen and Keltner (2017).

Clearly defining what is and is not an emotion word is difficult (e.g., Pavlenko, 2008) and may vary between populations, for example there may be differences in the emotion lexicons of older and younger adults (e.g., Schrauf & Sanchez, 2008). Therefore, after they completed the lexical decision experiment, participants were asked to rate on a 1-7 scale the extent to which the word stimuli referred to an emotion. We used Moseley et al.'s (2012) instructions with minor changes (e.g., the word 'speak' changed to 'describe' to be suitable to the written, rather than spoken stimuli used in this experiment). Emotion words had an average rating greater than five and the non-emotion words had average ratings less than three.

We also predicted that any emotion word effect may be influenced by the participants' knowledge of the word stimuli. Consequently, following the experimental tasks, participants provided a rating of familiarity for each word stimulus using methods drawn from Stadthagen-Gonzalez and Davis (2006). Words with a familiarity rating of '1' (indicating the participant did not know they word) were excluded from that participant's dataset (12 items across all included participants).

Word set matching was conducted using JASP software (JASP, 2020). Bayesian Independent T-Tests were used with a Bayes Factor (BF_{10}) criterion of less than 0.333, see Table 1. Only familiarity[1] and concreteness were outside this Bayes factor criterion and were therefore included as control variables in the statistical models. However, given that there may be concerns regarding multicollinearity between concreteness and imageability,

[1] Note that this was not the participants' familiarity ratings which were only used for the purposes of excluding items from the analysis. Familiarity ratings used in analyses were taken from Scott et al., 2019.

additional analyses were also run using a smaller set of 36 emotion and 36 non-emotion words for which concreteness matched and therefore was not required to be included as a control variable (Emotion word set: mean = 2.38, SD = 0.44; Non-emotion abstract word set: mean = 2.46, SD = 0.51; BF_{10} = 0.295).

[Table 1 near here]

To ensure that they were appropriate for the population from which our sample of participants were drawn, prior to the commencement of the study, ratings for arousal, valence, imageability and context availability were collected from another cohort of Macquarie University undergraduate students. Instructions for the ratings were drawn from previous studies (imageability: Stadthagen-Gonzalez & Davis, 2006; context availability: Schwanenflugel et al., 1988; valence and arousal: Scott et al., 2019). There were high correlations between our ratings for the word stimuli and existing arousal, valence and imageability norms (all > .80) (imageability: Bristol Norms, Stadthagen-Gonzalez & Davis, 2006; valence and arousal: Glasgow norms, Scott et al., 2019). Context availability ratings were not compared to existing norms due to a lack of pre-existing ratings for the stimuli.

Nonwords were generated from the English Lexicon Project database (Balota et al., 2007) and matched with real word stimuli for length (number of letters and number of phonemes), number of morphemes, number of syllables, mean bigram frequency and orthographic neighbourhood density.

Procedure

Lexical Decision Task. The experiment was run using PsychoPy software (Peirce et al., 2019). Prior to beginning the experiment, participants were provided with written and verbal instructions and completed 5 practice trials. The 312 experimental trials (52 emotion words, 52, non-emotion abstract words, 52 concrete word fillers, 156 nonwords) were presented in 3 blocks, with a 10 second break between blocks. Each 104-item block contained 52 words and 52 nonwords, including 17-18 each of emotion words, non-emotion abstract words and concrete filler words. Block order was counterbalanced between participants and order of trials was randomised. A trial began with a white fixation cross in the centre of a black screen. At 100ms, a lowercase letter string was displayed in the centre of the screen in white, size 60, Arial font. Participants provided a button press response and the trial ended when the participant responded, or after 2000ms if the participant did not respond.

Data Preparation.

Data Exclusion. Analysed data was limited to correct responses and responses with latencies greater than 200ms. Data was visually inspected for outliers and, in line with the recommendations of Baayen and Milin (2010), a conservative approach was taken with an aim of excluding minimal data (< 10%). Across all participants, there were 239 incorrect responses. Of the correct responses, one item was excluded due to a response time less than 200ms and 12 items excluded that were outside the main range on visual inspection. This represented 0.2% of the total correct datapoints with no more than two excluded outliers per participant.

Data Transformations. A Box-Cox test (Box & Cox, 1964) was run using the MASS package (Venables & Ripley, 2002) to determine which (if any) form of data transformation was appropriate. This indicated a log transformation was appropriate ($\boldsymbol{\lambda} \cong 0$). However, noting that concerns have been raised regarding the use of data transformation in linear mixed effects models (e.g., Balota et al., 2013; Cohen-Shikora et al., 2019; Wilcox & Rousselet, 2017), we followed the guidelines suggested by Brysbaert and Stevens (2018) and ran analyses on both transformed and untransformed data. We only interpreted effects as meaningful when they were significant in both analyses.

Variable Coding. Contrast coding was used for word type, task order and bilingual status. Continuous variables (imageability, context availability, arousal, extremity of valence, directional valence, familiarity, concreteness, presentation order) were standardised. As previous studies have commonly found that lexical processing is facilitated by both positive and negative valence (e.g., Kousta et al., 2009; Vinson et al., 2014), it was predicted that any effect of valence would be evident at both extremes of the scale (very positive and very negative). Consequently, to generate an extremity of valence measure, valence ratings were converted to a 1- 4 scale with neutral ratings low on the scale (0) and strongly valenced ratings (both positive and negative) high on the scale (4).

Analysis Methods.

Linear mixed effects modelling was used in two sets of analyses: first to examine for an effect of word type (Research Question 1.1) and second for an effect of word type, arousal, extremity of valence, imageability and context availability (Research Question 1.2). Analyses were run in R-studio (R Core Team, 2013) using the lme4 package (Bates et al., 2015). Details of the analysis parameters are provided in Table 2. For RQ 1.2, we used a hierarchical approach in which the order of variable inclusion was determined by those features most expected to influence outcomes based on previous modelling in Vinson et al. (2014) (i.e., extremity of valence and arousal, followed by imageability, followed by context availability which was not included in Vinson et al.). Extremity of valence and arousal were added simultaneously following Kousta et al. (2011) who expressed a preference for considering these variables together as 'affective associations' arguing that for abstract words there was no clear rationale for expecting dissociations between them. Planned comparisons of the models were carried out using Likelihood Ratio Tests (ANOVA). Following the recommendations of Matuschek et al. (2017) in relation to concerns regarding loss of power with more complex models, where there was no significant difference between models, we report the simpler model. In response to the recommendations of an anonymous reviewer, relating to the potential for Type I errors (Barr et al., 2013), we also ran maximal models with the most complete random effects structure that would converge. Any change in the model outcomes is reported in the relevant section of the Results, and maximal model outputs are provided in the Supplementary Materials (S3).

[Table 2 near here]

Results

Mean accuracy on the lexical decision task was close to ceiling at 96.6% (SD = 3.16). Therefore, we analysed only response time data (for correct responses).

Analysis 1.1: Effect of Word Type

Analysis using log transformed data showed only a marginal difference between models with and without by participant random slopes for word type (χ^2 (2) = 5.77, p = .056).

Therefore, we report the model without random slopes here (see Table 3 for a summary of LME output). Model outputs showed that word type (emotion vs. non-emotion abstract words) did not significantly predict variance in response times. This is reflected in the mean response times for each set (mean RT: Emotion = 625ms, SD = 170; Non-emotion abstract = 621ms, SD = 166). Comparison of this model and a base model without the word type experimental variable showed no significant difference between the models, further supporting that word type did not contribute significantly to the fit of the model ($\chi^2(1) = 0.004$, $p = .952$). Of the control variables, presentation order, response time on the previous trial and familiarity showed significant effects in the expected directions, while there was no significant effect of concreteness or task order.

As outlined above, we examined outcomes in analyses using both log transformed and untransformed data with an effect only considered meaningful if this was found across both sets of analyses (Brysbaert & Stevens, 2019). In contrast to the analysis using transformed data, analysis with untransformed data showed a significant difference between models with and without random slopes for word type ($\chi^2(2) = 10.99$, $p = .004$ and therefore the model with random slopes is reported (Table 3). The pattern of fixed and random effects results was unchanged from those found for the log transformed data: there was no significant effect of word type, adding word type to the base model did not improve the fit of the model and there were significant effects of the control variables presentation order, response time on the previous trial and familiarity.

[Table 3 near here]

Analysis 1.2: Effect of Extremity of Valence, Arousal, Imageability and Context Availability

In Analysis 1.2, we examined the effects of extremity of valence, arousal, imageability and context availability on the lexical decision response times. Comparisons indicated that there was only a significant difference in model fit between Model 1 (extremity of valence, arousal) and Model 2 (extremity of valence, arousal, imageability; Log transformed data: $\chi^2(1) = 5.608$, $p = .018$; Untransformed data: $\chi^2(1) = 6.761$, $p = .009$, see Table 1 for an outline of Model structures). Consequently, as Model 2 best fit the data, the outcomes of this model are reported in Table 4.

Interpretation of fixed effect estimates of the best fitting model indicated a significant effect of imageability, but not of the other experimental variables. Of the control variables, familiarity, presentation order and response time of previous trial showed significant effects in the expected directions, consistent with Analysis 1.1. These results were true of both log transformed and untransformed data.

[Table 4 near here]

Preregistered Exploratory Analyses

Additional planned exploratory analyses were conducted to determine whether the outcomes were impacted by multicollinearity between fixed effects (specifically, between the experimental variables and between imageability and concreteness). As a null-effect was found for extremity of valence, we also carried out (planned) exploratory analyses to examine whether this could be accounted for by direction of valence (negative or positive) or due to the effect being modulated by an interaction between valence and arousal or valence and word frequency. The results of these analyses are reported below.

Multicollinearity of Fixed Effects. Having closely correlated fixed effects impacts the ability to interpret the effect of any one variable (see for example, Yoo et al., 2014). In our analysis, the intercorrelations between experimental variables were relatively low (all $r < 0.5$). Moreover, the Variance Inflation Factors (VIFs) in all the models were less than 1.75 (see Table 4). This is below the cut-off of 2.5 which has been interpreted as indicating minimal concern for multicollinearity consistent with the recommendations of Allison (1999). Nonetheless, to err on the side of caution, we addressed the possible impacts of multicollinearity in two further ways: first the correlation between concreteness and imageability was addressed by analysing a smaller subset of words (36 emotion words and 36 non-emotion abstract words) which were matched on concreteness and therefore allowed concreteness to be removed from the model. Secondly, we ran single variable models to examine the effect of each variable on the model independently of the other experimental variables.

Concreteness-Matched Word Sets. Analyses 1.1 and 1.2 were repeated but using concreteness-matched word sets and without concreteness included as a fixed effect. These analyses showed two changes to the results using the larger word set: in the analysis with untransformed data there was no significant difference between models with and without random slopes in Analysis 1.1. All other results for both Analysis 1.1 and Analysis 1.2 showed the same pattern reported above with the larger stimulus set both with log transformed and untransformed data.

Models with a Single Experimental Variable. We also ran models in which the fixed effects included only one experimental variable (extremity of valence, arousal, imageability or context availability which were compared with a base model with no experimental variables. Model comparisons showed a significant difference between a model with arousal as the only experimental variable and the base model (Log transformed data: $\chi^2(1) = 4.974$, p = .026, Untransformed data: $\chi^2(1) = 4.583$, p = .032) and between the model with imageability and the base model (Log transformed data: $\chi^2(1) = 8.399$, p = .004, Untransformed data: $\chi^2(1) = 9.398$, p = .002). The single variable model including extremity of valence and the model including context availability did not show a significantly improved fit to the data when compared with the base model.

Direction of Valence. The main analysis of valence was conducted using a collapsed 1-4 scale (as described in the Methods) in order to capture the effect of extremity of valence. However, as noted previously, some researchers have reported the effect of valence to be directional, with a processing advantage for positive words over negative words (e.g., Estes & Aldeman, 2008; Kuperman et al., 2014, Larson et al., 2008; Sutton & Lutz, 2019). To determine if direction of valence impacted on outcomes, we also conducted an analysis using the same model structures as Analysis 1.2 but using continuous valence ratings (based on a 1-7 scale where 1 = negative, 7 = positive; standardised). This also showed no evidence of a significant effect for valence and the pattern of results was unchanged: Model structure 2 in which imageability was added to the model remained a significantly better fit than Model structure 1 (valence and arousal as the experimental variables) (Log transformed data: $\chi^2(1) = 5.827$ p = .017; Untransformed data: $\chi^2(1) = 6.903$, p = .009). All other model comparisons remained non-significant.

Interaction of Valence with Frequency and Arousal. It has also been suggested that valence may interact with frequency (Barriga-Paulino et al., 2022, Kuperman et al., 2014, Scott et al., 2014) and arousal (Barriga-Paulino et al., 2022, Citron et al., 2014, Larsen et al., 2008). We used log frequencies drawn from the HAL (Lund & Burgess, 1996) database to calculate the frequency interactions. The interactions were not significant when added to the best fitting model (Model 2) although the interaction between extremity of valence and frequency was marginal in the untransformed data analysis (extremity of valence*frequency: Log transformed data: β = 2.57e-03, SE = 1.84e-03, t(116.1) = -1.14, p = .161; Untransformed data: β = 2.16, SE = 1.30, t(113.9) = 1.66, p = .097.

When directional valence was used in place of the extremity of valence measure, the interaction between frequency and directional valence was marginal when both log transformed and untransformed data were used (Log Transformed data: β = 0.01, SE = 4.36e-03, t(92.4) = 1.77, p = .079; Untransformed data: β = 6.05, SE = 3.08, t(92.1) = 1.97, p = .052).

Discussion

The motivation for this experiment was to investigate whether there were differences in lexical decision response times between subsets of abstract words, specifically, abstract words that label emotions ('emotion words') and other abstract words ('non-emotion abstract words'). However, we found no evidence to support an effect of word type (emotion words/non-emotion abstract words). This was also in line with the results from Kousta et al. (2011) and Vinson et al. (2014). Vinson et al. showed that emotion words did not differ from other abstract words (in lexical decision response times), but instead argued that differences in performance were based on how emotional a word was, as measured by ratings of extremity of valence and arousal.

Our second analysis allowed us to further investigate how the outcomes may have been influenced by the semantic variables: extremity of valence, arousal, imageability, and context availability. Only imageability was shown to have a significant impact on performance. This is in contrast with Kousta et al.'s (2011) position that extremity of valence and arousal impact on processing of abstract words while imageability plays a greater role in the processing of concrete words (see also Vigliocco et al., 2014; Vinson et al., 2014). We also carried out (planned) exploratory analyses that ruled out any impact of multicollinearity on our effects: Analyses using word sets that were matched for concreteness, and models with single experimental variables produced results consistent with the outcomes of the more complex models. It was also not the case that the lack of an effect of valence was due to the impact of valence direction (positivity/negativity). Despite previous research which suggest faster processing times for positively valenced words compared with negatively valenced words (e.g., Estes & Aldeman, 2008; Kuperman et al., 2014, Sutton & Lutz, 2019), our analyses still showed no effect of valence when using a continuous scale. A marginal interaction was found between frequency and both measures of valence (extremity of valence and directional valence) suggesting that it may be that the valence effects were impacted by the frequency of our stimuli. However, this was only evident in the untransformed data analysis and so should be interpreted with caution.

In summary, the results of this experiment support an effect of imageability on processing of emotion and non-emotion abstract words in a lexical decision task but did not provide evidence supporting a difference based on word type (emotion/non-emotion abstract) or valence or context availability. In the second experiment, we examined if this pattern of results would be evident in a word production task: reading aloud.

Experiment 2: Reading Aloud

In this experiment, we examined if an effect of word type and/or of the experimental variables (extremity of valence, arousal, imageability, context availability) was evident in reading aloud. While lexical decision has been used extensively to examine the effects of emotion word type and affective semantic features (e.g., Altarriba & Bauer, 2004; Kousta et al., 2009; 2011; Kuperman et al., 2014; Vinson et al., 2014), there has been relatively little examination on the effects in production tasks in unimpaired populations. Estes and Aldeman (2008) and Larsen et al. (2008) report an effect of valence on reading aloud response times, however, neither of these studies examined the role of valence in emotion words specifically. Lexical decision and reading aloud utilise (partially) differing underlying lexical processes and previous studies have demonstrated differences in effects of semantic variables between these tasks. Of particular significance, in the context of the findings of Experiment 1, greater effects of imageability have been found for lexical decision than reading aloud (Balota et al., 2004; Ferrand et al., 2011). It was therefore of interest whether the effect of imageability would be maintained in our reading aloud experiment, as well as whether effects of word type (emotion vs. non-emotion abstract), valence and arousal or context availability would be apparent.

Methods

Participants

The participants were the same as in Experiment 1. Participants completed both experiments on the same day and the order of Experiment 1 and Experiment 2 was counterbalanced between participants. As in Experiment 1, we examined whether there was a significant effect of bilingualism. Bilingualism did not show a significant effect and a significant interaction between bilingualism and the experimental fixed effects was not evident. However, in order for participant sets to be consistent with the lexical decision experiment, we only included the monolingual data sets in the analysis. In addition, technical difficulties resulted in four datasets failing to record, these datasets were therefore absent from the analysis. In total, data from 63 participants was included in the analysis (Mean Age = 20.62 years, SD = 5.52; 42 Female, 18 Male, 3 Other).

Materials

The stimuli were the word stimuli used in Experiment 1.

Procedure

Reading Aloud Task. The reading aloud task was run using PsychoPy software (Peirce et al., 2019). Participants were presented with a single word in size 60 white Arial font on a black background and instructed to read it aloud as quickly as possible. Words were presented in randomised order. Each word stimulus was shown for two seconds. Voice recordings were initiated at the beginning of the trial and continued 1 second after the stimulus disappeared. Between each trial participants saw a black screen with a centred white fixation cross for 0.5 seconds. Voice recordings were collected using PsychoPy software and a compression microphone. Participants were instructed to only say the word on the screen and avoid use of fillers (e.g., ‘um’). Three practice trials were completed in which participants received feedback. Once the experimental trials commenced, participants were not provided with feedback on accuracy.

Data Preparation.

Response Time Calculations. Response times were manually encoded using PRAAT software (Versions 6.3.01; Boersma & Weenink, 2021). Incorrect responses, including responses with beginning with fillers (um, ah), self-corrections, comments or non-verbal output (e.g., coughing/sneezing) were excluded from analysis.

Three coders measured the response times: the first author and two research interns. The data from 18 participants (28% participant datasets) were first coded by all the coders, this resulted in 85% agreement within 10 ms and a Pearson's correlation of r = 0.97- 0.98 between each of the coders (the RT measurements of the first author were used for analysis of these data sets). At this point, the remaining datasets were divided between the three coders and completed independently.

Data Exclusion. Data exclusions followed the same procedure as Experiment 1. A total of 26 items were excluded representing 0.004% of the total correct responses across all included participants and with no more than 2 items excluded from any single participant's dataset.

Data Transformations. A Box Cox analysis (Box & Cox, 1964) indicated inverse transformation was appropriate.

Analysis Methods. Analysis followed the same procedure as outlined in Experiment 1 (see Table 2, earlier) with two analyses: Analysis 2.1 with the experimental variable of word type and Analysis 2.2 with hierarchical addition of experimental variables (Model 1: arousal and valence; Model 2: arousal, valence and imageability; Model 3: arousal, valence, imageability and context availability; Model 4: with the semantic variables of valence, arousal, imageability and context availability as fixed variables. Control variables were as in Experiment 1: familiarity, concreteness, presentation order, response time at previous (correct) trial and task order.

Results

Participants had near-ceiling accuracy of 96.4% (SD = 4.11) and therefore as in lexical decision, we focused our analysis on response times for correct responses.

Analysis 2.1: Effect of Word Type

In the first analysis, we examined whether word type (non-emotion abstract vs. emotion words) would significantly impact reading response times. Comparison between models with and without by participant random slopes for word type was not significant (Inverse transformed data: χ^2 (2) = 0.26, p = .880; Untransformed data: χ^2 (2) = 3.71, p = .156) indicating that there was no significant variance by participant on the word type variable. Therefore, the simpler model without by participant random slopes is reported in Table 5. Comparison of models with and without word type (but retaining all control variables) was significant (Inverse transformed data: χ^2 (1) = 5.05, p = .025; Untransformed data: χ^2 (1) = 4.16, p = .041) indicating that adding word type significantly improved the fit of the model. In support of this finding, examination of the fixed effect estimates indicated a significant effect of word type on response times and this was evident in modelling using both inverse transformed and untransformed data (see Table 5). The direction of this effect was in line with expectations with mean RTs for emotion words being faster than non-emotion abstract words (non-emotion: mean = 656ms, SD = 112; emotion: mean = 639ms, SD = 109). Of the control variables, there was no significant effect of concreteness or task order, but presentation order, response time on the previous trial and familiarity showed significant effects in the expected directions, consistent with the lexical decision results in Experiment 1.

In Analysis 2.2, we examined whether the effect of word type could be accounted for by the effects of any of the semantic variables of interest (imageability, context availability, valence and arousal) and whether any of those variables independently influenced reading aloud response times.

[Table 5 near here]

Analysis 2.2: Effect of Valence, Arousal, Imageability and Context Availability.

A hierarchical approach was used, beginning with the simplest model and building to the most complex (see Table 1, earlier). All model comparisons were non-significant. As in Analysis 2.1, the control variables of presentation order, response time on the previous trial and familiarity showed significant effects in the expected directions.

As we wished to investigate the effect of word type found in Analysis 2.1, we examined Model 4, which contained all experimental variables including word type (see Table 6). Model comparison between this model, and a model without word type (Model 3) was not significant, indicating that word type did not significantly influence reading latencies.

[Table 6 near here]

Preregistered Exploratory Analyses

Planned exploratory analyses were conducted as in Experiment 1, to determine whether the outcomes were impacted by multicollinearity between fixed effects, direction of valence (negative or positive) or an interaction between valence and frequency or arousal.

Multicollinearity of Fixed Effects. As in lexical decision, Variance Inflation Factors (VIFs) in all the models were low (< 1.53) and below the cut-off of 2.5 indicating minimal concern for multicollinearity (Allison, 1999) (see Table 6 for VIFs). As in Experiment 1, we used concreteness-matched word sets to address potential multicollinearity between concreteness and imageability. The results were consistent with those from the main analyses using the larger word sets.

We also ran models in which the fixed effects included only one experimental variable (valence, arousal, imageability or context availability). None of these variables showed significant effects in analyses with either inverse transformed or untransformed data, supporting the results from the more complex models.

Direction of Valence. Use of directional valence in place of extremity of valence did not change the pattern of outcomes.

Interaction of Valence with Arousal and Frequency. We also found no significant interactions between valence and arousal or frequency using either extremity of valence or directional valence measures in the best fitting model (Model 1, experimental variables: valence and arousal).

Discussion

This experiment aimed to investigate if emotion words are processed differently to non-emotion abstract words in reading aloud. We additionally explored the effects of valence, arousal, imageability and context availability. This is the first experiment, to our knowledge, to directly compare reading of emotion and non-emotion abstract words.

In contrast to the lexical decision task in Experiment 1, we found a significant effect of word type on response times in reading: emotion words were read aloud faster than non-emotion abstract words. This is consistent with previous research claiming emotion words have a processing advantage over other abstract words (e.g., Altarriba & Bauer, 2004). Yet, this effect could not be attributed to effects of any of the key semantic variables (imageability, context availability, valence, arousal). In Analysis 2.2, none of the model comparisons were significant, indicating that none of the variables improved the fit of the model. This included when we added word type to the model with the other experimental variables. In the absence of a significant effect of any of the semantic variables (valence, arousal, imageability and context availability), that the word type effect was not maintained in this analysis is difficult to interpret and may indicate that the word type effect was relatively fragile.

That we did not find any significant effects of the semantic variables is in line with previous research which suggests that reading aloud response times are more influenced by factors relating to lexical form (e.g., initial phoneme, word length) than semantic factors (e.g., Balota et al., 2004; Ferrand et al., 2011). However, as any difference in response times for emotion and non-emotion abstract words would be expected to be underpinned by semantics, the word type effect found in our first analysis is unexpected. This will be discussed in greater detail in the General Discussion.

General Discussion

This study aimed to further shed light on abstract word processing, with a particular focus on words that describe emotions. Our experiments built on the existing literature regarding emotion word effects in lexical decision and added new data regarding reading aloud. We aimed to address whether there was evidence that emotion words were processed differently compared to other 'non-emotion' abstract words and, if so, whether the underlying mechanism of this effect could be attributed to other psycholinguistic properties. For example, as reported in the Introduction, Altarriba and Bauer (2004) reported that emotion words were better recalled than non-emotion abstract words. They attributed this pattern to higher imageability and context availability ratings in emotion words compared to non-emotion abstract words. However, their word sets were relatively small (N = 30) and were only matched on frequency and word length, presenting the possibility that other lexical variables (e.g., age of acquisition, familiarity) may have influenced the emotion word effect. Similarly, Kousta et al. (2011) suggested that Altarriba and Bauer's findings may have been influenced by the inherent emotional characteristics of emotion words and correspondingly strong ratings in valence and arousal. In a previous study, Kousta et al. (2009) had shown that valence had a significant effect (and arousal had a marginal effect) on response time in lexical decision across all words with emotional associations. Kousta et al. (2009) included robust control of a wide range of variables ((log) frequency, number of letters, concreteness, imageability, familiarity, age of acquisition, orthographic neighbourhood density, number of morphemes, mean positional bigram frequency). However, they did not specifically examine a set of words which label emotions. Consequently, we aimed to explore this area of research further by investigating whether emotion and non-emotion abstract words showed differences in processing when large word sets closely matched for a large range of psycholinguistic variables were used. Specifically, we matched sets for the same variables as Kousta et al. (2009) as well as additional variables which may impact on lexical decision and/or reading aloud: phonological neighbourhood density (e.g., Yates et al., 2004), Orthographic Levenshtein Distance (e.g., Ferrand et al., 2011; Yarkoni et al., 2008), Phonological Levenshtein Distance (Ferrand et al., 2011). We also explored whether an effect of imageability, context availability, valence and arousal influenced word processing, and if an effect of word type was found, whether these variables underpinned this effect.

In lexical decision we found no evidence to support a difference in processing between emotion and non-emotion abstract words, while in reading aloud a difference between the word types was evident but only in a model that included this as the sole experimental variable. Surprisingly, we also found no evidence for effects of valence, arousal or context availability on either lexical decision or reading aloud performance. While imageability did show a significant effect, it was only in lexical decision. We will first focus our discussion on the semantic variables and then turn to the effect of emotion words as a subtype of abstract words.

Effects of Semantic Variables: Imageability, Context Availability, Arousal and Valence

Our finding of an effect of imageability in lexical decision and not reading aloud is largely in line with previous research. For example, Balota et al. (2004) showed that semantic variables, including imageability, were stronger predictors of latency for lexical decision than reading aloud (see also, Cortese & Khanna, 2007; Dymarska et al., 2023; Ferrand et al., 2011). Our study adds further evidence to this pattern and shows that the pattern is maintained in words which are exclusively abstract and with relatively low imageability ratings (mean = 3.42, SD = 0.83). A discussion of the underlying causes of the difference between lexical decision and reading aloud, and the implications for how we can interpret our findings across the four semantic variables focused on in this study (imageability, context availability, valence and arousal), will be presented at the end of this section.

Context availability did not show any significant effects in either experiment, even when it was the only experimental variable in the model. We included this variable based on the fact that Schwanenflugel et al (1983) argued that concrete words are of higher context availability than abstract words and that this contributed to the faster response times for concrete words in lexical decision. In addition, Altarriba and Bauer (2004) extended this hypothesis arguing that as emotion words are higher in context availability than other abstract words, this may contribute to their superior recall in the free recall task. However, while they found emotion words have higher context availability ratings than non-emotion abstract words, Altarriba and Bauer did not test the effects of context availability on free recall experimentally. We do not find any support for an advantage for emotion words in lexical decision or reading aloud overall. This this may be due to task effects (see discussion below). We also have some concerns regarding the reliability of the measurement of this variable. We took a rigorous approach to ensuring the quality of our ratings by collecting ratings from a relatively large sample of participants (*N* = 49 compared to *N* = 22 for Schwanenflugel et al., and *N* = ~8/9 for Altarriba & Bauer). Additionally, we excluded participant sets which had weak or no correlation to the mean of the other raters ($r^2 < 0.1$), following data trimming methods described by Brysbaert et al. (2014). However, when collating the ratings, we noted that, compared to other variables (valence, arousal and imageability), more participant sets were excluded for context availability (from 49 sets in total: 6 sets for context availability, 2 for arousal and imageability and 0 for valence).This may suggest that the raters had more difficulty understanding how to rate words on this variable and the ratings may not be reliable and/or valid.

Of greater theoretical interest was that we found little evidence for effects of either of the affective variables, valence and arousal, in the best fitting model on either task. This contrasts with our expectations based on the important role that these variables have been hypothesised to have in the processing and representation of abstract words (Kousta et al., 2011; Vigliocco et al, 2014). We will address each of these variables independently.

Arousal was not a significant predictor of response times in the best fitting model for either lexical decision or reading aloud. However, analysis using single experimental variable models did indicate that a model with only arousal was a better fit than the base model with no experimental variables. Given that evidence for an effect of arousal from the literature has been mixed, the weak effect found in our study might not be surprising. For example, while Kuperman et al. (2014) showed that low arousal words had faster response times than high arousal words in lexical decision, Kever et al. (2019) found the opposite effect. In addition, other authors have found a limited effect (e.g., Kousta et al., 2009 - marginal effect; Vinson et al., 2014 - significant effect only when arousal was the sole experimental variable). In sum, our study adds to the overall pattern that arousal appears to have minimal and inconsistent effects on abstract word processing.

Valence did not show a significant effect on performance on either task, whether as a measure of strength of valence (high valence = strongly positively or negatively valenced; low valence = neutral valence), or when it was measured as a directional continuous measure (low = negatively valenced; mid = neutral valenced; high = positively valenced). As previous studies have suggested valence may show an interaction with arousal (Barriga-Paulino et al., 2022; Citron et al., 2014; Larsen et al., 2008) or with frequency (Barriga-Paulino et al., 2022; Kuperman et al., 2014; Scott et al., 2014), we also examined these interactions. However, our analysis did not show any evidence that the frequency or arousal characteristics of the stimuli influenced the (lack of) effects of valence.

It is clear therefore that our study provides no support for an effect of valence on performance for either lexical decision or reading aloud. This contrasts with the position of Kousta et al. (2011) and Vigliocco et al. (2014) who, finding a significant effect in lexical decision, assign a particularly important role to valence in the processing of abstract words. Within an embodied approach to lexical representations, they suggest that for abstract and introspective concepts, valence and arousal provide sensorimotor grounding. Vigliocco et al. further found a significant interaction between concreteness and valence, demonstrating that the strength of valence showed a greater effect on lexical decision response times in low concreteness (abstract) words than in high concreteness words. While most research examining effects of valence has concentrated on lexical decision (e.g., Kousta et al., 2009; 2011, Kuperman et al., 2014, Vigliocco et al, 2014, Vinson et al., 2014), Estes and Aldeman

(2008) and Larsen et al. (2008) showed an effect of valence in both lexical decision and reading aloud, although this was across a wider range of stimuli that were not limited to abstract words. Therefore, the lack of an effect of valence in either of our experiments contradicts expectations based on the published literature.

In summary, across both experimental tasks, we found limited to no significant effects of the semantic variables: imageability, context availability, arousal and valence. While, for each variable, there may be independent reasons for the outcomes, the weak effects overall may point to issues with assumption that the lexical decision and reading aloud tasks can reliably show semantic effects. We chose these tasks as they have been used extensively to examine the effects of semantics on word processing in previous studies (e.g., Balota et al., 2004; Cortese & Khanna, 2007; Yap et al., 2011) and therefore provided a substantial research body for comparison with our results. Yet, neither of these tasks *require* semantics; lexical decision hinges on access to the orthographic form and reading aloud requires access to the orthographic form and corresponding phonological form (e.g., Coltheart & Rastle, 1994; Seidenberg & McClelland, 1989). Therefore, to explore why our results showed limited effects, we can consider how semantics is thought to influence processing on these specific tasks. Hino and Lupker (1996) propose the role of semantics in activation of the orthographic form occurs via a feedback system. The authors showed that polysemous words had faster response times in lexical decision than words with single meanings. They interpreted this effect as showing that there is feedback from the semantic system during activation of the word-form resulting in words associated with richer semantic information (such as words with multiple meanings) receiving a processing advantage over words associated with less semantic information. Focusing specifically on lexical decision, Hino and Lupker noted that both high and low frequency words with multiple meanings have a processing advantage in in lexical decision, but that the facilitative effect from rich semantic representations is limited to low frequency words in reading aloud (see also Strain et al., 1995; Strain & Herdman, 1999 for further on the limited effects of semantics on high frequency words in reading aloud). This could be a potential explanation for the differences in semantic effects in lexical decision and reading aloud that we found for imageability as it would predict that imageability effects may be muted in reading aloud compared to lexical decision in word sets containing (some) high frequency words as was the case for our stimuli (HAL raw frequency range 304 – 165 830). However, this hypothesis is not supported by our data given that we did not find a significant interaction between frequency and (the semantic factors) valence or arousal in reading aloud.

While feedback from semantics may influence activation of the word-form, we should also consider the strength of these effects. Studies comparing predictors of lexical decision and reading aloud response times have shown that, while significant, the impact of semantic factors appears to be weaker than other factors. For example, Ferrand et al (2011) showed weak correlations between lexical decision response time and imageability compared to lexical variables (age of acquisition, frequency; for similar findings see also Balota et al, 2004; Cortese & Khanna, 2007). Therefore, while semantic factors may impact on response times in lexical decision and reading aloud, these effects are likely the relatively small and, consequently may not reach significance.

Effects of Abstract Word Type: Emotion vs Non-emotion

We were interested in the emotion word effect reported in the literature, and differences in semantic attributes (reflected in effects of semantic variables) as a potential mechanism underpinning this proposed processing advantage for emotion words over non-emotion abstract words. We first looked at whether this processing advantage was apparent in lexical decision and/or reading aloud in a stimulus set where emotion and non-emotion abstract words were closely matched for many important variables including concreteness. We then tested whether any effect of this variable was robust when the focus semantic variables discussed in the above section (valence, arousal, imageability and context availability) were additionally controlled. We found no evidence for an effect of word type in lexical decision but a significant effect in reading aloud. However, this effect was not maintained when the additional semantic variables were added to the model, even though none of these variables significantly predicted performance themselves.

The semantic variables of valence, arousal, imageability and context availability were selected as points of focus due to their relevance to key accounts of the emotion word effect. Altarriba and Bauer (2004) suggested that the emotion word effect stems from higher ratings of imageability and context availability in emotion words over non-emotion abstract words, while in contrast, Kousta et al. (2011) argued that the observed processing advantage of emotion words in Altarriba and Bauer (free recall) resulted from the influence of valence and arousal. Kousta et al. (2009; 2011) found that valence influenced lexical decision response times for abstract words, and Vinson et al. (2014) additionally demonstrated that lexical decision for both emotion and non-emotion words was influenced by valence. As the only significant effect of a semantic variable in our main analyses, was of imageability in lexical decision, one might have expected that any discernible emotion word effect would also be found in lexical decision rather than reading aloud. However, if anything, we found the reverse, and neither of the prior mentioned theoretical positions (imageability/context availability or valence/arousal) appear to explain the word type effect found in reading aloud.

Alternative views have been proposed regarding whether valence and arousal underpin the 'meaning' of emotion words which could also explain the null effects for these two semantic dimensions in our study. Fontaine et al. (2007; 2013; 2022) and Gillioz et al. (2016) hypothesised that, in addition to valence and arousal, the dimensions of 'novelty' and 'dominance' were needed to fully capture the 'meaning' of an emotion word. A model containing all four of these dimensions could therefore provide a more accurate predictor of emotion word processing. This is possible, however, Cowen et al. (2019; Cowen & Kelter, 2017) questioned how reliably arousal and valence map onto emotion words. Cowen et al. presented participants with emotional stimuli (non-lexical vocalisations) and asked them to assign an emotional 'category' and affective ratings (including arousal and valence) to each stimulus. While emotional category could explain the majority of the variance in valence and arousal ratings, the opposite was not true: arousal and valence ratings did not reliably predict emotion category. Consequently, arousal and valence might not be expected to strongly predict lexical processing of emotion words, which would be consistent with the null effects for these dimensions found in our study.

Returning the role of imageability in distinguishing emotion and non-emotion abstract words, it should be noted that although not matched, our word sets were very close in imageability (Emotion Set: Mean = 3.63, SD = 0.72; Non-emotion abstract emotion: Mean = 3.22, SD = 0.89; Bayes Factor_{10} = 3.53). Therefore, it is possible that this explains why a word type difference was not found, even in lexical decision. If this account were true, then it would support a role of imageability in distinguishing emotion and non-emotion words, however we cannot confirm this with our stimuli.

Summary and Conclusions

In summary, our study showed emotion words did not have a processing advantage over other abstract words in lexical decision and the effect in reading aloud was not robust when other semantic variables were added to the model. For the semantic variables, our finding of an imageability effect in lexical decision but not reading aloud is consistent with previous findings. However, the other variables did not show a significant effect on performance in either task (with the exception of arousal in a single variable model in lexical decision), and there was no observable relationship between the effects of the semantic variables and word type (emotion vs. non-emotion abstract). Our findings could perhaps also illustrate the possible limitations of lexical decision and reading aloud tasks as a means of testing the influence of semantic factors in lexical processing: the influence of semantic factors on these tasks may be too small to be reliably observed. Future studies may therefore consider effects in other tasks which have a greater requirement for access to semantic information such as semantic decision, synonym judgments or speeded categorisation. In addition, examining processing of emotion words and the effects of valence and arousal in people with aphasia may be fruitful. Previous studies have shown that people with aphasia may have specific difficulties with abstract word processing (e.g., Sandberg & Kiran, 2014) and therefore larger effects of factors relating to abstract words may be evident with this population.

Acknowledgements

We extend our thanks to Serje Robidoux for his immense help with statistical approaches and code and to Maria Korochkina for her assistance and advice regarding power analysis and preregistration. Two anonymous reviewers gave helpful suggestions on an earlier version of the manuscript.

Declaration of Interest

The authors have no conflict of interests to declare. This research was supported by an Australian Research Council discovery project grant (DP190101490) to the last author (LN) and a Research Training Program (RTP) scholarship (allocation number: 20191031) to the first author (CM).

Data Availability Statement

Data, code, preregistration documents and supplementary materials can be accessed at: https://osf.io/tyjse/files/osfstorage

References

Allison, P.D. (1999). Multiple regression: a primer. Pine Forge Press.

Altarriba, J., & Basnight-Brown, D. M. (2011). The representation of emotion vs. emotion-laden words in English and Spanish in the Affective Simon Task. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *15*(3), 310–328. https://doi.org/10.1177/1367006910379261

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology*, *117*(3), 389–410. https://doi.org/10.2307/4149007

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, *31*, 578-602. https://doi.org/10.3758/BF03200738

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28. https://doi.org/10.21500/20112084.807

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive Effects of Word Frequency and Stimulus Quality: The Influence of Trial History and Data Transformations. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(5), 1563–1571. https://doi.org/10.1037/a0032186

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology. General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barriga-Paulino, C. I., Guerreiro, M., Faísca, L., & Reis, A. (2022). Does emotional valence modulate word recognition? A behavioral study manipulating frequency and arousal. *Acta Psychologica*, *223*, 103484–103484. https://doi.org/10.1016/j.actpsy.2021.103484

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67* (1), 1-48. https://doi.org/jss.v067.i01.

Boersma, P. & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.3.01.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26* (2), 211-243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2* (1), 16–16. https://doi.org/10.5334/joc.72

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1* (1), 9–9. https://doi.org/10.5334/joc.10

Brysbaert, M., Warriner, A.B. & Kuperman, V. (2014) Concreteness ratings for 40 thousand generally known English word lemmas. *Behavioral Research, 46*, 904–911 https://doi.org/10.3758/s13428-013-0403-5

Catricalà, E., Della Rosa, P. A., Plebani, V., Vigliocco, G., & Cappa, S. F. (2014). Abstract and concrete categories? Evidences from neurodegenerative diseases. *Neuropsychologia*, *64*, 271–281. https://doi.org/10.1016/j.neuropsychologia.2014.09.041

Citron, F. M. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, *56* (Apr), 79–89. https://doi.org/10.1016/j.neuropsychologia.2014.01.002

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45* (9), 1703–1723. https://doi.org/10.1037/xlm0000670

Colombo, L., & Burani, C. (2002). The influence of age of acquisition, root frequency, and context availability in processing nouns and verbs. *Brain and Language*, *81* (1–3), 398–411. https://doi.org/10.1006/brln.2001.2533

Coltheart, M. & Rastle, K. (1994). Serial processing in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 20* (6), 1197-1211.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology (2006)*, *60* (8), 1072–1082. https://doi.org/10.1080/17470210701315467

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist, 74*(6), 698–712. https://doi.org/10.1037/amp0000399

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the national academy of sciences*, *114* (38), E7900-E7909. https://doi.org/10.1073/pnas.1702247114

Crossfield, E., & Damian, M. F. (2021). The role of valence in word processing: Evidence from lexical decision and emotional Stroop tasks. *Acta Psychologica*, *218*, 103359–103359. https://doi.org/10.1016/j.actpsy.2021.103359

Dymarska, A., Connell, L., & Banks, B. (2023). Weaker than you might imagine: Determining imageability effects on word recognition. *Journal of Memory and Language*, *129*, 104398-. https://doi.org/10.1016/j.jml.2022.104398

Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion, 8*(4), 441–444. https://doi.org/10.1037/1528-3542.8.4.441

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: evidence from chronolex. *Frontiers in Psychology*, *2*, 306–306. https://doi.org/10.3389/fpsyg.2011.00306

Fontaine, J. R., Gillioz, C., Soriano, C., & Scherer, K. R. (2022). Linear and non-linear relationships among the dimensions representing the cognitive structure of emotion. *Cognition and Emotion*, *36*(3), 411-432. https://doi.org/10.1080/02699931.2021.2013163

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science*, *18*(12), 1050-1057. https://doi.org/10.1111/j.1467-9280.2007.020

Fontaine, J. R. J., Scherer, K. R., & Soriano, C. (2013). *Components of emotional meaning: A sourcebook*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199592746.001.0001

Gillioz, C., Fontaine, J. R., Soriano, C., & Scherer, K. R. (2016). Mapping emotion terms into affective space. *Swiss Journal of Psychology*, *75*(3), 141–148.https://doi.org/10.1024/1421-0185/a000180

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1331–1356. https://doi.org/10.1037/0096-1523.22.6.1331

Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2020). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, *35*(7), 813–839. https://doi.org/10.1080/23273798.2019.1620957

Hsieh, S., Foxe, D., Leslie, F., Savage, S., Piguet, O., & Hodges, J. R. (2012). Grief and joy: Emotion word comprehension in the dementias. *Neuropsychology*, *26*(5), 624–630. https://doi.org/10.1037/a0029326

JASP Team (2020). JASP (Version 0.17.3)[Computer software]. https://jasp-stats.org/

Kazanas, S. A., & Altarriba, J. (2015). The automatic activation of emotion and emotion-laden words: Evidence from a masked and unmasked priming paradigm. *The American journal of psychology*, *128*(3), 323-336. https://doi.org/10.5406/amerjpsyc.128.3.0323

Kever, A., Grynberg, D., Szmalec, A., Smalle, E., & Vermeulen, N. (2019). "Passion" versus "patience": the effects of valence and arousal on constructive word recognition. *Cognition and Emotion*, *33*(6), 1302–1309. https://doi.org/10.1080/02699931.2018.1561419

Knickerbocker, H., & Altarriba, J. (2013). Differential repetition blindness with emotion and emotion-laden word types. *Visual Cognition*, *21*(5), 599–627. https://doi.org/10.1080/13506285.2013.815297

Knickerbocker, H., Johnson, R. L., & Altarriba, J. (2015). Emotion effects during reading: Influence of an emotion target word on eye movements and processing. *Cognition and Emotion*, *29*(5), 784–806. https://doi.org/10.1080/02699931.2014.938023

Kousta, S.T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology. General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kousta, S.T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology. General*, *143*(3), 1065–1081. https://doi.org/10.1037/a0035669

Landis, T., Graves, R., & Goodglass, H. (1982). Aphasic reading and writing: Possible evidence for right hemisphere participation. *Cortex*, *18*(1), 105–112. https://doi.org/10.1016/S0010-9452(82)80022-1

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*(4), 445–452. https://doi.org/10.1037/1528-3542.8.4.445

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers*, *28*(2), 203-208. https://doi.org/10.3758/bf03204766

Lund, T. C., Sidhu, D. M., & Pexman, P. M. (2019). Sensitivity to emotion information in children's lexical processing. *Cognition*, *190*, 61–71. https://doi.org/10.1016/j.cognition.2019.04.017

Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language*, *19*(1), 124–141. https://doi.org/10.1016/0093-934X(83)90059-7

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language*, *94*, 305-315. https://doi.org/10.1016/j.jml.2017.01.001

Moseley, R., Carota, F., Hauk, O., Mohr, B., & Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex, 22*(7), 1634–1647. https://doi.org/10.1093/cercor/bhr238

Newton, C., Thornley, H., & Bruce, C. (2020). The influence of emotional valence on word recognition in people with aphasia. *Language, Cognition and Neuroscience*, *35*(8), 1064–1072. https://doi.org/10.1080/23273798.2020.1713385

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Canadian Journal of Psychology*, *45*(Sep 91), 255–287.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1p2), 1–25. https://doi.org/10.1037/h0025327

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism*, *11*(2), 147–164. https://doi.org/10.1017/S1366728908003283

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pulvermüller, F. (2018). The case of CAUSE: neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences*, *373*(1752), 20170129–20170129. https://doi.org/10.1098/rstb.2017.0129

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Sandberg, C., & Kiran, S. (2014). Analysis of abstract and concrete word processing in persons with aphasia and age-matched neurologically healthy adults using fMRI. *Neurocase*, *20*(4), 361–388. https://doi.org/10.1080/13554794.2013.770881

Schrauf, R. W., & Sanchez, J. (2008). Using Freelisting to Identify, Assess, and Characterize Age Differences in Shared Cultural Domains. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *63*(6), S385–S393. https://doi.org/10.1093/geronb/63.6.S385

Schwanenflugel, P. J., Akin, C., & Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104. https://doi.org/10.3758/BF03208259

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499–520. https://doi.org/10.1016/0749-596X(88)90022-8

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, *24*(1), 114–126. https://doi.org/10.2307/748013

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2014). Emotion words and categories: evidence from lexical decision. *Cognitive Processing*, *15*(2), 209–215. https://doi.org/10.1007/s10339-013-0589-6

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96* (4), 523-568.

Souter, N. E., Lindquist, K. A., & Jefferies, E. (2021). Impaired emotion perception and categorization in semantic aphasia. *Neuropsychologia*, *162*, 108052–108052. https://doi.org/10.1016/j.neuropsychologia.2021.108052

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605. https://doi.org/10.3758/BF03193891

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*(4), 347–359. https://doi.org/10.1037/h0087322

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1140–1154. https://doi.org/10.1037/0278-7393.21.5.1140

Sutton, T. M., & Altarriba, J. (2008). Emotion words in the mental lexicon: A new look at the emotional Stroop effect. *The Mental Lexicon*, *3*(1), 29–46. https://doi.org/10.1075/ml.3.1.04sut

Sutton, T. M., & Lutz, C. (2019). Attentional Capture for Emotional Words and Images: The Importance of Valence and Arousal. *Canadian Journal of Experimental Psychology*, *73*(1), 47–54. https://doi.org/10.1037/cep0000154

Venables, W. N. & Ripley, B. D. (2002) Modern applied statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: the role of emotion. *Cerebral Cortex, 24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vinson, D., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition and Emotion*, *28*(4), 737–746. https://doi.org/10.1080/02699931.2013.851068

Wang, X., Shangguan, C., & Lu, J. (2019). Time course of emotion effects during emotion-label and emotion-laden word processing. *Neuroscience Letters*, *699*, 1–7. https://doi.org/10.1016/j.neulet.2019.01.028

Wilcox, R. R., & Rousselet, G. A. (2017). A guide to robust statistical methods in neuroscience. *Current Protocols in Neuroscience, 82*:8.42.1–8.42.8. https://doi.org/10.1101/151811

Wu, C., & Zhang, J. (2019). Conflict processing is modulated by positive emotion word type in second language: An ERP study. *Journal of Psycholinguistic Research*, *48*(5), 1203–1216. https://doi.org/10.1007/s10936-019-09653-y

Wu, C., Zhang, J., & Yuan, Z. (2021). Can masked emotion-laden words prime emotion-label words? An ERP test on the mediated account. *Frontiers in Psychology*, *12*, 721783–721783. https://doi.org/10.3389/fpsyg.2021.721783

Yap M. J., Tan S. E., Pexman P. M., & Hargreaves I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, 18(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y

Yap, M. J., & Seow, C. S. (2014). The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic Bulletin & Review*, *21*(2), 526–533. https://doi.org/10.3758/s13423-013-0525-x

Yarkoni, T., Balota, D. & Yap, M. (2008) Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. https://doi.org/10.3758/PBR.15.5.971

Yates, M., Locker, L. & Simpson, G.B. (2004) The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11,* 452–457. https://doi.org/10.3758/BF03196594

Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., & Lillard, J. W., Jr (2014). A study of effects of multicollinearity in the multivariable Analysis. *International journal of applied science and technology*, *4*(5), 9–19.

Zhang, J., Wu, C., Yuan, Z., & Meng, Y. (2019). Differentiating emotion-label words and emotion-laden words in emotion conflict: an ERP study. *Experimental Brain Research*, *237*(9), 2423–2430. https://doi.org/10.1007/s00221-019-05600-4

Table 1

Word Set Matching

	Independent Samples T-Test (df = 102)			Bayesian Independent Samples T-Test	
	t	df	p	BF_{10}	error %
Arousal	-3.235	102	.002	19.237	2.488e-6
Valence	-10.041	102	< .001	6.139e +13	4.201e -17
Imageability	-2.543	102	.012	3.529	7.819e -6
Context availability	0.038	102	.969	0.207	0.022
Age of acquisition	0.927	102	.356	0.304	0.031
Concreteness	4.285	102	< .001	480.111	8.367e -6
Familiarity	-1.124	102	.264	0.364	0.033
Log freq Subtlex	0.852	102	.396	0.286	0.030
Log freq HAL	0.165	102	.869	0.21	0.022
No. syllables	0.640	102	.524	0.249	0.027
No. letters	-0.760	102	.449	0.268	0.029
No. phonemes	-0.364	102	.716	0.220	0.024
No. morphemes	-0.271	102	.787	0.214	0.023
OLD	-0.897	102	.372	0.297	0.031
PLD	-0.452	102	.652	0.227	0.025
Orthographic N	-0.986	102	.326	0.320	0.032
Phonological N	-0.804	102	.423	0.276	0.029
Mean bigram freq	0.640	102	.524	0.249	0.027

Notes: Log freq Subtlex = Log spoken frequency from the Subtlex database; Log freq HAL = Log written frequency from the HAL database; OLD = Orthographic Levenshtein Distance; PLD = Phonological Levenshtein Distance; Orthographic N = Orthographic neighbourhood; Phonological N = Phonological neighbourhood; Mean bigram freq = Mean bigram frequency; t = t-value, df = degrees of freedom; p = p-value; BF_{10} = Bayes Factor. Norms for concreteness, log frequency (HAL), log frequency (Subtlex), orthographic Levenshtein distance, phonological Levenshtein distance, orthographic neighbourhood, phonological neighbourhood, number of letters, number of morphemes, number of syllables and number of phonemes are from the English Lexicon Project (Balota et al., 2007). Familiarity norms from Scott et al., (2019).

Commented [LN1]: I think it would be good to have the mean and SD for each variable for each set here too. You can delete the df column

Table 2

Linear Mixed Effects Models for Experiments 1 and 2

	Analysis 1.1 (RQ1.1)	Analysis 1.2 (RQ1.2)
Sampling Units	6212 observations; 104 items (52 non-emotion abstract words; 52 emotion words); 67 participants	As in Analysis 1.1
Dependent Variable	Response time for correct responses (both transformed or untransformed)	As in Analysis 1.1
Experimental Variables	Word type	*Inclusion in the following order: 1.Valence & Arousal 2. Imageability 3. Context availability 4. Word type
Control Variables	**Familiarity, concreteness, response time at previous correct trial, presentation order, task order	As in Analysis 1.1
Random Effects	Random intercepts for item and participant	Random intercepts for item and participant.
Model Comparisons (LRT)	Comparison with a model with by-participant random slopes for word type	Model 1 vs Model 2 (inclusion of imageability) Model 2 vs. Model 3 (inclusion of context availability) Model 3 vs. Model 4 (inclusion of word type)

Notes. LRT = Likelihood Ratio Tests. *The control variables of task order and familiarity, were omitted from the preregistration document in error and Model 2 was erroneously described as adding word category rather than imageability. **Age of acquisition was removed as it was matched between the word sets.

Table 3

Linear Mixed Effect model output examining the effects of word type (emotion vs non-emotion abstract words) on lexical decision response times for transformed and untransformed data for Analysis 1.1

	Log Transformed Data Analysis						**Untransformed Data Analysis**					
Fixed Effects												
	Estimates	*SE*	*CI*	*t*	*p*	*VIF*	*Estimates*	*SE*	*CI*	*t*	*p*	*VIF*
Intercept	6.39	0.02	6.35 – 6.43	311.65	<.001		615.69	13.65	588.94 – 642.44	45.12	**<.001**	
Experimental Variable												
Word Type	3.11e-04	0.01	-0.01 – 0.01	0.06	.953	1.195	0.49	3.97	-7.30 – 8.28	0.12	.902	1.170
Control Variables												
Familiarity	-0.04	5.00e-03	-0.05 – -0.03	-7.81	-0.04	1.014	-26.95	3.54	-33.88 – -20.02	-7.62	-26.95	1.012
Concreteness	-0.01	0.01	-0.03 – 0.01	-0.66	-0.01	1.180	-3.68	7.96	-19.28 – 11.92	-0.46	-3.68	1.157
Pres Order	-4.52e-03	9.66e-04	-0.01 – -0.00	-4.68	-4.52e-03	1.008	-2.75	0.71	-4.13 – -1.36	-3.88	-2.75	1.008
RT Previous Trial	0.03	2.14e-03	0.03 – 0.04	14.63	0.03	1.008	21.01	1.57	17.93 – 24.08	13.39	21.01	1.008
Task Order	0.01	0.02	-0.02 – 0.05	0.65	0.01	1.000	5.03	12.26	-19.00 – 29.06	0.41	5.03	1.000
Random Effects												
σ2	0.03						15440.92					
τ00 item	0.002						962.99					
τ00 participant	0.02						9896.17					
τ11 participant.1.word set							115.12					
ρ01 participant.1							-0.06					
ICC	0.46						0.42					
Function	Log(RT) ~ Word Type + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)						RT ~ Word Type + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 + Word Type | Participant)					
Sampling Units	6212 observations; 104 items; 67 participants						6212 observations; 104 items; 67 participants					

Notes. RT = Response Time. Pres Order = Presentation Order; RT Previous Trial = Response time of previous trial, SE = Standard Error; CI = Confidence interval; VIF = Variance Inflation Factor, σ2 = residual variance, τ00 item= random variance, τ11 = random slope variance, ρ01 = random intercept/slope correlation, ICC = intraclass correlation coefficient

Table 4

Linear Mixed Effect model output for Analysis 1.2, Model 2 examining the effects of arousal, valence, imageability on lexical decision response times.

Fixed Effects	**Log Transformed Data Analysis**						**Untransformed Data Analysis**					
	Estimate	*SE*	*CI*	*t-value*	*p-value*	*VIF*	*Estimate*	*SE*	*CI*	*t-value*	*p-value*	*VIF*
Intercept	6.38	0.02	6.34 – 6.43	304.46	**<.001**		610.56	14.00	583.12 – 637.99	43.62	**<.001**	
Experimental Variables												
Arousal	-7.120e-03	0.01	-0.02 – 0.00	-1.19	.233	1.699	-4.17	4.23	-12.45 – 4.12	-0.99	.324	1.702
Valence	-3.468e-04	2.00e-03	-0.004 - 0.004	-0.17	.863	1.389	-0.49	1.42	-3.28 – 2.30	-0.34	.733	1.392
Imageability	-0.02	0.01	-0.05 – -0.00	-2.34	**.019**	1.552	-19.37	7.51	-34.09 – -4.64	-2.58	**.010**	1.553
Control Variables												
Familiarity	-0.03	0.01	-0.04 – -0.02	-5.69	**<.001**	1.284	-20.93	3.81	-28.40 – -13.45	-5.49	**<.001**	1.283
Concreteness	1.119e-03	0.01	-0.02 – 0.02	0.10	.918	1.206	2.96	7.71	-12.16 – 18.07	0.38	.701	1.207
Pres Order	-4.53e-03	9.66e-04	-0.01 – -0.00	-4.69	**<.001**	1.008	-2.76	0.71	-4.15 – -1.37	-3.89	**<.001**	1.008
RT Previous Trial	0.03	2.14e-03	0.03 – 0.04	14.63	**<.001**	1.008	20.98	1.57	17.91 – 24.06	13.36	**<.001**	1.008
Task Order	0.01	0.02	-0.02 – 0.05	0.65	.518	1.000	5.1	12.28	-18.96 – 29.17	0.42	.678	1.000
Random Effects												
σ2	0.03						15556.39					
τ00 item	0.002						860.59					
τ00 participant	0.02						9912.03					
ICC	0.46						0.41					
Function	Log(RT) ~ Arousal + Valence + Imageability + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)						RT ~ Arousal + Valence + Imageability + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)					
Sampling Units	6212 observations; 104 items; 67 participants						6212 observations; 104 items; 67 participants					

Notes. RT = Response Time. Pres Order = Presentation Order; RT Previous Trial = Response time of previous trial, SE = Standard Error; CI = Confidence interval; VIF = Variance Inflation Factor, σ2 = residual variance, τ00 item= random variance, ICC = intraclass correlation coefficient

Table 5

Linear Mixed Effect model output for Analysis 2.1 examining the effects of word type (emotion vs non-emotion abstract words) on reading aloud response times.

	Inverse Transformed Data Analysis						Untransformed Data Analysis					
Fixed Effects												
	Estimate	*SE*	*CI*	*t-value*	*p-value*	*VIF*	*Estimate*	*SE*	*CI*	*t-value*	*p-value*	*VIF*
Intercept	1.59e-03	2.62e-05	1.54e-03 - 1.64e-03	60.81	**<.001**		645.04	11.17	623.14 – 666.95	57.73	**<.001**	
Experimental Variable												
Word Type	1.67e-05	7.46e-06	2.18e-06 - 3.13e-05	2.24	**.025**	1.195	-6.49	3.20	-12.75 – -0.22	-2.03	**.042**	1.195
Control Variables												
Familiarity	4.56e-05	7.03e-06	3.19e-05 - 5.93e-05	6.49	**<.001**	1.013	-19.56	3.01	-25.46 – -13.66	-6.50	**<.001**	1.013
Concreteness	1.08e-05	1.58e-05	-2.00e-05 - 4.17e-05	0.68	.494	1.181	-3.89	6.78	-17.18 – 9.40	-0.57	.566	1.181
Pres Order	4.93e-06	1.97e-06	1.06e-06 – 8.80e-06	2.50	**.013**	1.001	-0.89	0.89	-2.63 – 0.85	-1.00	.318	1.001
RT Previous Trial	-3.47e-05	1.96e-06	-3.85e-05 - -3.08e-05	-17.71	**<.001**	1.001	14.71	0.88	12.99 – 16.43	16.73	**<.001**	1.001
Task Order	-3.50e-06	2.32e-05	-4.89e-05 - 4.20e-05	-0.15	.880	1.000	2.12	9.89	-17.26 – 21.50	0.21	.830	1.000
Random Effects												
σ2		2.29e-08					4625.3					
τ00 item		4.46e-09					809.77					
τ00 participant		3.36e-08					6110.09					
ICC		0.62					0.60					
Function	1/RT ~ Word Type + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)						RT ~ Word Type + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)					
Sampling Units	6101 observations; 104 items; 63 participants						6101 observations; 104 items; 63 participants					

Notes. RT = Response Time. Pres Order = Presentation Order; RT Previous Trial = Response time of previous trial, SE = Standard Error; CI = Confidence interval ; VIF = Variance Inflation Factor, σ2 = residual variance, τ00 item= random variance, ICC = intraclass correlation coefficient

Table 6

Linear Mixed Effect model output for Analysis 1.2, Model 4 examining the effects of word type (emotion/non-emotion abstract), arousal, valence, imageability, context availability on reading aloud response times.

	Inverse Transformed Data Analysis						**Untransformed Data Analysis**					
Fixed Effects												
	Estimate	*SE*	*CI*	*t-value*	*p-value*	*VIF*	*Estimate*	*SE*	*CI*	*t-value*	*p-value*	*VIF*
Intercept	1.60e-03	2.75e-05	1.54e-03 - 1.65e-03	58.13	**<.001**		643.97	11.71	621.02 – 666.93	54.98	**<.001**	
Experimental Variables												
Arousal	2.27e-06	1.01e-05	-1.69e-05 - 2.15e-05	0.22	.822	2.146	-1.36	4.3	-9.79 – 7.07	-0.32	.752	2.145
Valence	-9.02e-07	3.53e-06	-7.64e-06 - 5.84e-06	-0.26	.799	1.830	0.32	1.51	-2.64 – 3.28	0.21	.832	1.830
Imageability	6.38e-06	1.68e-05	-2.56e-05 - 3.84e-05	0.38	.704	1.728	-3.13	7.17	-17.18 – 10.92	-0.44	.662	1.727
Context Av	-3.10e-06	7.27e-06	-1.70e-05 - 1.08e-05	-0.43	.671	1.098	1.6	3.11	-4.50 – 7.69	0.51	.608	1.098
Word Type	1.38e-05	9.68e-06	-4.68e-06 - 3.23e-05	1.43	.157	1.936	-5.07	4.14	-13.19 – 3.05	-1.22	.221	1.936
Control Variables												
Familiarity	4.47e-05	8.18e-06	2.91e-05 - 6.03e-05	5.47	**<.001**	1.322	-19.00	3.5	-25.86 – -12.14	-5.43	-19	1.322
Concreteness	5.90e-06	1.89e-05	-3.02e-05 - 4.20e-05	0.31	.756	1.620	-1.51	8.08	-17.35 – 14.34	-0.19	-1.51	1.620
Pres Order	4.92e-06	1.97e-06	1.05e-06 – 8.79e-06	2.49	.013	1.001	-0.88	0.89	-2.62 – 0.86	-0.99	-0.88	1.001
RT Previous Trial	-3.47e-05	1.96e-06	-3.85e-05 - -3.08e-05	-17.71	**<.001**	1.001	14.71	0.88	12.98 – 16.43	16.72	14.71	1.001
Task Order	-3.50e-06	2.32e-05	-4.89e-05 - 4.20e-05	-0.15	.881	1.000	2.12	9.89	-17.26 – 21.50	0.21	2.12	1.000
Random Effects												
σ2	2.29e-08						4625.27					
τ00 item	4.64e-09						842.19					
τ00 participant	3.36e-08						6110.19					
ICC	0.63						0.60					
Function	1/RT ~ Word Type + Arousal + Valence + Context Av. +Imageability + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)						RT ~ Word Type + Arousal + Valence + Context Av. +Imageability + Familiarity + Concreteness + Pres Order + RT Previous Trial + Task Order + (1 | Item) + (1 | Participant)					
Sampling Units	6101 observations; 104 items; 63 participants						6101 observations; 104 items; 63 participants					

Notes. RT = Response Time. Pres Order = Presentation Order; RT Previous Trial = Response time of previous trial, SE = Standard Error; CI = Confidence interval ; VIF = Variance Inflation Factor, σ2 = residual variance, τ00 item= random variance, ICC = intraclass correlation coefficient.