

Arousal and emotional valence interact in written word recognition

Francesca M.M. Citron^{1,2}, Brendan S. Weekes^{1,3} & Evelyn C. Ferstl^{1,4}

¹School of Psychology, University of Sussex, UK

²Cluster of Excellence “Languages of Emotion”, Freie Universität Berlin, Germany

³Laboratory for Communication Science, The University of Hong Kong

⁴Center for Cognitive Science, Albert-Ludwigs-Universität Freiburg, Germany

Address correspondence to: Francesca Citron
Cluster of Excellence “Languages of Emotion”
Freie Universität Berlin
Habelschwerdter Allee 45
D-14195 Berlin

Tel.: +49 (0) 30 838 57869
Fax: +49 (0) 30 838 52887
E-mail: fmm.citron@gmail.com

Running head: Arousal and valence interaction

Key words: word recognition, valence, arousal, emotion, lexical decision, gender

Abstract

Behavioural, psychophysiological and neuroimaging studies reveal a prioritisation for emotional material in a variety of cognitive tasks. Although emotion is comprised of two dimensions (valence and arousal), previous research using verbal materials has mostly focused on valence, while controlling level of arousal. The aim of the present study is to investigate the effects of valence and arousal on lexical decision (LD) by manipulating both dimensions while controlling correlated psycholinguistic variables (e.g., word length, frequency, imageability). Results showed that valence and arousal affect word recognition in an interactive way: LD latencies are slower for positive high-arousal and negative low-arousal words compared to positive low-arousal and negative high-arousal words, in line with an approach-withdrawal tendency model. Furthermore, principal component analysis (PCA) on the latencies revealed a unique contribution of a distinct cluster of emotion variables, independent of lexico-semantic variables, to explaining written word recognition. We conclude that the dimensions of valence and arousal both need to be taken into account in studies of emotion word processing as they have an interactive relationship.

Acknowledgements

FMMC would like to thank Cristina Burani for her advice on the data analysis. This work is part of FMMC's doctoral dissertation, funded by a Graduate Teaching Assistant scholarship from the University of Sussex, UK. Parts of this study were presented at the conferences Architecture and Mechanisms of Language Processing in 2010 as well as at the meetings of the British Psychology Society and the Society for Psychophysiological Research in 2009.

Introduction

Emotion is generally characterised as a two-dimensional construct: emotional valence describes the extent to which an emotion is positive or negative whereas emotional arousal refers to its intensity i.e., whether an emotion is exciting/agitating or calming/sedating (Feldman Barrett & Russell, 1998). Typically, more valenced (positive or negative) stimuli are also more arousing (Bradley & Lang, 1999) and stimuli with a negative valence are rated higher in arousal than stimuli with a positive valence (e.g., Citron, Weekes, & Ferstl, in press). We know emotion has an impact on cognitive and language processing. For example, the emotional content of verbal material affects behavioural performance in a variety of tasks (e.g., Algom, Chajut, & Lev, 2004; Larsen, Mercer, & Balota, 2006; Nasrallah, Carmel, & Lavie, 2009), and this is reflected in psychophysiological and hemodynamic brain activity (Citron, 2012).

Effects of emotional valence on written word processing

A large body of research investigating the contribution of emotion to written word processing has focused on valence. For example, in the emotional Stroop task participants are required to name the colour of the font in which negative, neutral and (sometimes) positive words are written. Slower responding to negative words is usually observed (e.g., Algom et al., 2004; Kahan & Hely, 2008) and this effect has been interpreted as support for the automatic vigilance hypothesis (Pratto & John, 1991). The automatic vigilance hypothesis assumes that negative stimuli capture and hold attention due to their potentially threatening nature. Therefore, fewer resources are available for the cognitive task at hand and performance will be slower for items with a negative valence compared to positive or neutral stimuli. Slower responding to

negative words has also been reported in lexical decision (LD) tasks (Algom et al., 2004; Estes & Verges, 2008) and more errors are observed in response to negative words in emotional judgement tasks (emotional vs. neutral, Nasrallah et al., 2009).

One methodological problem with prior studies reporting effects of emotional valence and arousal on written word recognition is that dimensions of emotion on one hand, and lexico-semantic variables, such as word frequency, imageability and age of acquisition, on the other, are not independent but often correlated. If putative effects of emotion variables are merely a consequence of multi-collinearity with such variables, then the theoretical importance of emotion effects is dubious. Larsen et al. (2006) addressed this issue by re-analysing data from 32 studies investigating the emotional Stroop effect and showed that, after controlling for differences between conditions in correlated variables (word frequency, length and orthographic neighbourhood measures), the assumed processing cost for negative words disappeared (see also Kousta, Vinson, & Vigliocco, 2009 for the control of additional variables). Furthermore, few previous studies have investigated effects of emotional valence by controlling for the level of arousal as well as psycholinguistic variables and by presenting emotional and neutral words in a randomly mixed order. Interestingly, Algom et al. (2004) reported that slower responding to negative words disappears with random presentation, which prevents carryover effects between emotional words, previously obtained with block designs (e.g., Nasrallah et al., 2009). Studies using well controlled stimuli, randomised presentation and tasks that do not require any explicit evaluation of the emotional content of stimuli (LD, structural or semantic judgement tasks) show no difference in performance between positive and negative words, although such words are identified more quickly and accurately when

compared to neutral words (e.g., Kanske & Kotz, 2007; Kousta et al., 2009; Schacht & Sommer, 2009; Scott, O'Donnell, Leuthold, & Sereno, 2009).

The processing advantage obtained for emotional words over neutral words - henceforth called the emotionality effect - can be reconciled with emotional Stroop effects showing slower responses for negative than positive words if we consider the results of McKenna and Sharma (2004). They showed that a pseudo randomisation of trials in emotional Stroop tasks reversed the direction of the emotionality effect seen using block designs. However, it is not known if emotionality interacts with arousal when variables are manipulated in a fully randomised experimental design using an implicit processing task (the LD task).

How do arousal and valence affect written word processing?

Robinson, Storbeck, Meier and Kirkeby (2004) proposed a model of emotion processing with assumptions based on the combined effects of emotional valence and arousal. According to this model, stimuli with negative valence (e.g., fake) or with high arousal (e.g., a loud noise) elicit a withdrawal orientation because they represent a possible threat, whereas stimuli with positive valence (e.g., cake) or low arousal (e.g., a newsletter) elicit approach because they are perceived as safe. Robinson et al. assume that these two orientations are initiated independently at a pre-attentive level and are subsequently integrated to evaluate stimuli for further action. According to this model, positive low-arousal and negative high-arousal stimuli should be easier to process because they elicit congruent orientations toward an object in the environment (approach and withdrawal, respectively), whereas difficulty of integration will arise for positive high-arousal and negative low-arousal stimuli, which elicit conflicting approach-withdrawal tendencies. Robinson et al. provide empirical support for their

assumptions: in a series of experiments using emotional pictures, written words and tasks requiring explicit emotional evaluation, behavioural responses were slower if stimuli elicited conflicting tendencies. Robinson et al.'s model thus allows interactive effects of arousal and valence on written word recognition and indeed predicts such an interaction with greatest processing costs for positive high-arousal and negative low-arousal stimuli.

Recently, a few studies have further investigated the effects of both emotional variables on word processing. Evidence of an interaction between valence and arousal in word recognition was reported by Larsen, Mercer, Balota and Strube (2008). In their study, arousal modulated speed of access to negative word representations. Specifically, LD performance for negative low-arousal words was slower than negative high-arousal and positive (high or low-arousal) words. Hofmann, Kuchinke, Tamm, Võ and Jacobs (2009) similarly report faster LD for negative high-arousal words compared to negative low-arousal and neutral words, and faster LD to positive low-arousal than neutral words. Further, Bayer, Sommer and Schacht (2012) report slower LD latencies to negative low-arousal words compared to negative high-arousal as well as positive low-arousal and neutral low-arousal words. However, despite the expectation derived from the Robinson et al. model, not all studies find an interaction between emotion dimensions in written word recognition. Some studies report that emotional valence is a stronger predictor than emotional arousal of LD performance and naming latency (Estes & Adelman, 2008; Kousta et al., 2009).

The present study

Our first aim was to explore the effects of valence and arousal on single word processing and to test predictions from Robinson's model by manipulating both

variables in a LD task, while controlling over correlated lexico-semantic variables. In contrast to Larsen et al. (2008) and Kousta et al. (2009), we selected stimuli from our own dataset (Citron et al., in press) based on homogeneous rating instructions i.e., using exactly the same procedures for rating affective variables and lexico-semantic properties. Our second aim was to determine if emotion constrains LD performance beyond other word properties. We expected affective word properties to predict LD performance, beyond other lexico-semantic properties. Furthermore, we predicted faster LD for valenced (negative and positive) words compared to neutral words as well as an interaction between emotional valence and arousal, with slower LD in conditions eliciting conflicting approach-withdrawal orientations.

Methods

Participants

Forty-three native speakers of English from the University of Sussex (25 women, age range: 19-36 years, $M = 23.63$, $SD = 4.89$) performed a LD task. Participants were right-handed, with normal or corrected-to-normal vision, and had had no learning disability. Volunteers were either given course credits or paid £5.

Materials

One hundred and fifty words were selected from the SAWL norms (Citron et al., in press), containing ratings for critical affective features - emotional valence, arousal - and lexico-semantic features - familiarity, age of acquisition (AoA), imageability. Ratings were collected using a 7-point Likert scales: valence ranged from -3 (very negative) to +3 (very positive); arousal, familiarity and imageability were scaled from 1 (not at all) to 7 (very high); for AoA, age ranges in years were

given: 0-2, 2-4, 4-6, 6-9, 9-12, 12-16, older than 16, subsequently re-coded in 1-to-7 points. Length in letters, phonemes and syllables, frequency of use (spoken and written) as well as orthographic neighbourhood size (N-size) and orthographic neighbourhood frequency (N-frequency) values were taken from other databases (please refer to Citron et al., in press).

To determine the effects of emotion variables, the constructs “Emotionality” will refer to the comparison of positive, negative and neutral words and “Valence” will refer to the comparison of positive and negative stimuli only. Emotionality was manipulated by selecting 50 positive ($M = 1.74$, $SD = 0.36$), 50 negative ($M = -1.51$, $SD = 0.34$) and 50 neutral words ($M = 0.23$, $SD = 0.42$), matched for all length measures, N-size, N-frequency, word frequency (Log10), rated AoA and imageability ($F_s(2,147) < 1.66$). Words spanned all grammatical categories (nouns, adjectives, verbs and mixed categories). Each condition contained emotion-denoting words (e.g., *happy*, *sad*) as well as other valenced words (e.g., *flower*, *rain*). Positive and negative words were matched for arousal ($t(98) = -.98$, *ns*) and were both higher in arousal than neutral words (See Table 1 for descriptive statistics). As familiarity ratings were biased towards positive words in Citron et al.’s (in press) norms, this variable was not considered for the design of experimental conditions, but was included in additional analyses to exclude this possible bias (please refer to the Data Analysis Section).

Arousal was manipulated within valenced words, with half of the positive and negative words being high, and half low in arousal ($F(1,96) = 168.19$, $p < .001$), resulting in 4 conditions. The mean arousal level of high-arousal words was matched between positive and negative valenced words, as well as the mean arousal level of low-arousal words ($F(1,96) = 0.09$, *ns*). Nevertheless, high-arousal valenced words were more positive or negative than low-arousal valenced words ($t_s(48) > 2.79$ $p_s <$

.008). Stimuli could be matched for all lexical properties described above ($F_s(3,96) < 1.57$), except rated imageability $F(3,96) = 18.39, p < .001$, which was higher for high-arousal words. This was expected because arousal and imageability are highly correlated (see Citron et al., in press). Due to the potential impact of this unavoidable confound, the indirect effect of imageability was controlled in all analyses of valence by arousal only.

One hundred and fifty non-words were selected from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). Stimuli length ranged between 4-10 letters and 3-8 phonemes. More than half the items (53%) were word-like (i.e., they follow the orthographic and phonological rules of English). Words and non-words were matched for number of letters $t(289.22) = 1.51, ns$ and phonemes $t(298) = 0.55, ns$.

(Table 1 about here)

Procedure

The experiment was programmed with E-Prime software. Participants were seated in front of a computer monitor at a distance of approximately 70 cm. The stimuli were presented in the centre of the screen in non-capitalized white letters on a black background (24-point Courier font). Two letters subtended 1° of visual angle.

Participants were required to read letter strings and to decide whether they were English words or not, as accurately and as quickly as possible. A response box with two buttons corresponding to “yes/no” answers was provided and configuration was counterbalanced across participants. At the start of each trial, a fixation cross appeared in the centre for 800 ms, followed by a letter string, which remained until

participants made the LD. No time limit for the response was given. The screen was then blank for 1000 ms; after that a new trial would start.

A 10-trial practice block was followed by 6 experimental blocks, each one divided in 2 sessions, with a short break in between; 150 words and non-words were intermixed. Each block contained 25 words and 25 non-words, and an almost equal amount of positive, negative and neutral words. Block order and word order within blocks were randomised across participants. Reaction times (RTs) and accuracy to each item were recorded. The experiment lasted approximately 30 minutes.

Data analysis

Effects of emotionality and valence by arousal. For each participant, outlying RTs exceeding ± 3 *SDs* above the participant's mean RT, as well as incorrect trials, were excluded from the analysis. Mean RTs, mean accuracy rates and *SDs* for each participant and each condition, as well as for each stimulus, were calculated. As a standard procedure in psycholinguistic research, we performed all inferential statistical analyses by participant and by item, in order to consider both sources of variability (Clark, 1973). The results of the analyses by item should confirm those obtained in the analyses by participant and allow generalisation of the findings on the specific word sample to a broader set of words (i.e., the language lexicon). Nevertheless, given the large number of variables that influence word recognition (length, frequency, imageability, etc.), item analyses tend to show less significant or weaker effects than the participant analyses. Discrepancies between participant and item analyses will index non-robust effects. On the other hand, confirmation of the findings through careful control for possibly confounding variables and replication within subgroups of participants will strengthen the reliability of the findings.

The following analyses by participant (indexed by a subscripted 1) and by item (subscripted 2) were conducted: *t*-tests comparing words and non-words; ANOVAs on the 150 words with factor Emotionality (neutral, positive, negative) and on the 100 valenced words with factors Valence (positive, negative) x Arousal (high, low). In the latter design, imageability was controlled in the analysis by participant by first regressing raw RTs and accuracy rates for each participant on to item imageability ratings, and then by using the resulting standardised residuals as the dependent variable; in the analysis by item, imageability was used as a covariate.

Imageability was not controlled in the emotionality design because it was successfully matched in the manipulation; in this way, we avoided unnecessary removal of variance from the data.

Indirect effects of familiarity and self-reference. Citron et al. (in press) reported that positive words tend to be rated as more familiar overall and attributed this correlation to a response bias, similar to the one reported by Lewis, Critchley, Rotshtein and Dolan (2007) in a self-referential task: participants categorised positive words more often as referring to themselves than negative words. On the other hand, the correlation between familiarity and valence might be due to the specific stimulus selection in the norming study (Citron et al., in press). In order to explore this issue, we decided to control familiarity and self-reference, using fresh ratings for the latter feature¹. Ratings for both variables were compared across emotionality conditions and a correlation was computed. In order to investigate whether the pattern of results holds after removal of the apparent bias, the analyses of emotionality were repeated

¹ 69 participants (64 women), aged 18-34 years ($M = 19.64$; $SD = 2.19$) were instructed to rate “how much does each of the following words describe yourself” on a scale from 1 (not at all) to 7 (very much). All 150 words and some fillers were rated.

by partialling out the effects of familiarity and self-reference, using the procedure described for excluding the effects of imageability. Analyses of valence by arousal were also repeated by partialling out potential effects of familiarity and self-reference, along with imageability.

Post-hoc analyses of gender differences in emotion processing. Several studies have reported gender differences in emotion processing including enhanced processing of negative pictures by women when compared to men (Kemp, Silberstein, Armstrong, & Nathan, 2004), greater accuracy and sensitivity in labelling emotional facial expressions (Montagne, Kessels, Frigerio, de Haan, & Perrett, 2005), as well as higher accuracy and stronger BOLD response in detecting the congruency between emotional speech prosody and word valence (Schirmer, Zysset, Kotz, & von Cramon, 2004). Other studies report no gender differences in the ratings for affective content of written words (Bauer & Altarriba, 2008). To date, no study has investigated the effect of gender on written emotion word processing. Therefore, we repeated the analyses by participant described above by including the between-subjects factor Gender and we computed RTs and accuracy rates for female and male groups separately performing analyses by items again. Given evidence that women are usually more risk-avoidant than men (see Eckel & Grossman, 2002 for a review) and generally more sensitive to emotional information, we wondered whether the pattern of results will differ for men and women.

Contribution of emotional variables to predicting LD latencies. Given that measures of length, as well as emotionality and arousal, were highly correlated with each other ($r_s > 0.75$), a principal component analysis (PCA) was first performed on

all 300 words using our norms and 50 additional items² to extract the latent factors from all lexico-semantic and affective variables and to minimise any collinearity in the subsequent multiple regression analysis. A stepwise multiple regression analysis was conducted to investigate whether affective features contribute to predicting the mean LD latency for each word, beyond lexico-semantic features.

Results

Mean LD accuracy overall was high $M = 97%$, *standard error (SE) = 0.3%*. Words were recognised significantly faster ($t_1(42) = 3.47, p = .001$; $t_2(257.97)^3 = 12.64, p < .0001$) and more accurately ($t_1(42) = 2.54, p = .015$; $t_2(223.67) = 4.13, p < .0001$) than non-words (see Table 2a).

Emotionality

Analysis of RTs showed a main effect of emotionality in the participant analysis only ($F_1(2,84) = 5.92, p = .004$; $F_2(2,147) = 2.09, ns$). Pair-wise comparisons revealed faster RTs for valenced words compared to neutral words **in both analyses** ($F_1(1,42) = 9.21, p = .004$; $t_2(147) = 1.98, p = .049$), but no difference between positive and negative words ($F_1(1,42) = 1.46, ns$; $t_2(147) = 0.50, ns$). There was also a significant effect of emotionality on accuracy ($F_1(2,84) = 9.27, p < .0001$; $F_2(2,147) = 4.36, p = .014$), with more correct responses to positive words compared to negative and neutral words ($F_1(1,42) = 20.34, p < .0001$; $t_2(147) = 2.94, p = .004$) and no difference between negative and neutral words ($F_1(1,42) = 0.05, ns$; $t_2(147) = 0.24, ns$; see Table 2a).

(Table 2 about here)

² Ratings for 50 more items were additionally collected and then included in the PCA because increasing the number of items is beneficial for this analysis.

³ t-values and degrees of freedom for non-homogeneous variance are reported.

Emotional valence and arousal

In the imageability-corrected ANOVAs⁴, no significant main effects of valence or arousal on the residuals of the RTs were found (both $F_{s1}(1,42) < 0.30, ns$; $F_{s2}(1,95) < 0.24, ns$). A significant interaction between valence and arousal (cf. Figure 1a) showed larger absolute mean residual values in response to positive high-arousal (PH) and negative high-arousal (NL) words compared to positive low-arousal (PL) and negative high-arousal (NH) words in the participant analysis only, indicative of slower RTs for the former conditions ($F_I(1,42) = 6.05, p = .018$; $F_2(1,95) = 0.89, ns$).

Accuracy results showed an arousal effect in the participant analysis only ($F_I(1,42) = 13.18, p = .001$; $F_2(1,95) = 0.40, ns$), with a larger absolute mean residual value for high-arousal words, indicating higher accuracy. A valence effect was found in the item analysis only ($F_I(1,42) = 1.22, ns$; $F_2(1,95) = 5.68, p = .019$), with higher accuracy for positive words (see Table 2a). No interaction was found ($F_I(1,42) = 1.72, ns$; $F_2(1,95) = 0.53, ns$).

Indirect effects of familiarity and self-reference

We found that positive words were significantly more familiar than negative and neutral words ($t(147) = 4.33, p < .0001$; see Table 2b). Furthermore, self-

⁴ When imageability is not controlled in the analyses, a main effect of arousal (confounded with imageability) is obtained in the RTs ($F_I(1,42) = 29.41, p < .001$; $F_2(1,99) = 8.33, p = .005$), whereby highly arousing and imageable words are responded to faster, in line with the typical imageability effect (e.g., Bird, Franklin, & Howard, 2001; Paivio, Yuille, & Madigan, 1968). Please refer to Table 2a for the descriptive statistics. A significant interaction is found in the analysis by participant only ($F_I(1,42) = 5.75, p = .021$; $F_2(1,99) = 1.25, ns$) whereby the arousal/imageability effect is larger for negative words. No effect of valence is observed. Accuracy rates show significantly higher accuracy for positive words ($F_I(1,42) = 16.08, p < .001$; $F_2(1,99) = 6.15, p = .015$) and for highly arousing/imageable words only in the analysis by participant ($F_I(1,42) = 5.07, p = .030$; $F_2(1,99) = 2.21, ns$), but no interaction ($F_I(1,42) = 1.02, ns$; $F_2(1,99) = 0.43, ns$). Overall, these results show typical imageability effects and no apparent influence of imageability on valence, as previously reported by Kanske and Kotz (2007). We therefore conclude that the control of this variable in the main analyses helped us to partial out possibly confounding effects without hiding any interesting combined effects between emotional variables and imageability.

reference was correlated with familiarity ($r = 0.62, p < .0001$): positive words were rated as significantly more self-referential than negative and neutral words ($t(147) = 10.81, p < .0001$).

After partialling out effects of these variables, no main effect of emotionality was found in the mean residuals of both RTs and accuracy rates (both $F_{S_1}(2,84) < 0.74, ns$; $F_{S_2}(2,145) < 0.82, ns$). In the valence by arousal design, no main effects of valence or arousal on the mean residuals of the RTs were found (both $F_{S_1}(1,42) < 0.41, ns$; $F_{S_2}(2,145) < 1.49, ns$). Importantly, a significant interaction between valence and arousal was found in both participant and item analyses ($F_1(1,42) = 9.04, p = .004$; $F_2(1,93) = 4.36, p = .039$): absolute mean residuals of the RTs were larger, i.e., RTs were slower, to PH and NL words than to PL and NH words (see Figure 1a). The mean residuals of the accuracy rates confirmed a significant effect of arousal only in the participant analysis ($F_1(1,42) = 8.73, p = .005$; $F_2(1,93) = 0.35, ns$) and no other effects or interactions (both $F_{S_1}(1,42) < 2.06, ns$; $F_{S_2}(1,93) < 2.53, ns$).

(Figure 1 about here)

Post-hoc analyses of gender differences in emotion processing

Emotionality by gender. The main effects of emotionality on RTs ($F_1(2,82) = 6.16, p = .003$) and accuracy rates ($F_1(2,82) = 8.77, p < .0001$) were confirmed in the analyses by participant, i.e., valenced words were responded to faster than neutral ones ($F_1(1,41) = 9.01, p = .005$) and positive words more accurately than negative and neutral ones ($F_1(1,41) = 19.14, p < .0001$). No main effects of gender ($F_{S_1}(1,41) < 0.13, ns$) and no interactions between emotionality and gender ($F_{S_1}(2,82) < 0.92, ns$) were found. Finally, the item analyses for the two gender groups revealed only a trend

toward an emotionality effect in the male RTs in the direction mentioned above ($F_2(2,147) = 2.56, p = .081$; all other $F_{S_2}(2,147) < 1.52, ns$).

As in previous analyses, the effects of emotionality on RTs and accuracy disappeared after partialling out the effects of familiarity and self-reference (both $F_{S_1}(2,82) < 0.88, ns$). Further, no effects of gender (both $F_{S_1}(1,41) < 0.12, ns$) or interactions ($F_{S_1}(2,82) < 0.49, ns$) were found. Finally, in the analysis by item no significant effects were found within either gender (all $F_{S_2}(2,147) < 1.75, ns$).

Valence by arousal by gender. In the imageability-corrected ANOVA, analysis by participant, the significant interaction between valence and arousal on the residuals of the RTs was confirmed ($F_1(1,41) = 5.31, p = .026$): PH and NL words were responded to more slowly than PL and NH words. No main effect of valence, arousal or gender was found (all $F_{S_1}(1,41) < 0.31, ns$) and no other interactions were significant ($F_{S_1}(1,41) < 1.06, ns$). The main effect of arousal on the residuals of the accuracy rates was also confirmed in the analysis by participant ($F_1(1,41) = 12.02, p = .001$), along with no effect of valence ($F_1(1,41) = 1.90, ns$). Further, no main effect of gender ($F_1(1,41) = 0.08, ns$) and no interactions were found (all $F_{S_1}(1,41) < 2.65, ns$). When comparing gender groups, no main effects and no interaction were observed in either RTs or accuracy rates (all $F_{S_2}(1,95) < 1.56, ns$).

After partialling out the effects of familiarity, self-reference and imageability, the significant interaction between valence and arousal on the residuals of the RTs was once again confirmed ($F_1(1,41) = 7.94, p = .007$); the same interactive pattern was observed for female and male participants separately (see Figure 1b and 1c), but did not reach standard levels of significance (Women: $F_2(1,93) = 3.54, p = .063$; Men: $F_2(1,93) = 2.79, p = .098$), apparently because of loss of statistical power. No main effects of valence, arousal or gender ($F_{S_1}(1,41) < 0.24, ns$) and no other

significant interactions were found ($F_{S_I}(1,41) < 1.24, ns$). The mean residuals of the accuracy rates confirmed a significant effect of arousal in the participant analysis ($F_I(1,41) = 7.92, p = .007$), no effect of gender ($F_I(1,41) = 0.11, ns$) and no other effects or interactions (all $F_{S_I}(1,41) < 2.50, ns$). No significant effects were found within either gender (all $F_{S_2}(1,93) < 2.14, ns$).

Contribution of emotional variables to predicting LD latencies

In order to minimise collinearity among predictors of LD latency, 11 variables were entered into a PCA: emotionality, arousal, familiarity, AoA, imageability, log-frequency, length in letters, phonemes and syllables, N-size, N-frequency⁵. Four factors with eigenvalues > 0.9 emerged, as shown in Table 3 (a, b). All length and neighbourhood measures loaded on the first factor, which accounted for 42.1% of the variance. Log-frequency, familiarity and AoA loaded on the second factor (16.4% of variance). The two emotion measures loaded on the third factor (12.8% of variance). Finally, imageability loaded heavily on the fourth factor, together with AoA and length in phonemes (8.3% of variance).

These results show that affective variables cluster together and are distinct from lexico-semantic variables. Not surprisingly, AoA loaded on both lexical and semantic factors with approximately equal weight; this variable is in fact considered partly lexical and partly semantic (c.f. Juhasz, 2005).

In the multiple regression analysis, lexico-semantic factors were entered as predictors in a first step, followed by the emotion factor. They were all significant predictors. As shown in Table 3c, the lexical factors accounted for an equal amount of variance, followed by smaller contributions of semantic and emotional factors.

(Table 3 about here)

⁵ Self-reference could not be included in the PCA as ratings for this variable were collected post-hoc only for the 150 items used in the study. Therefore, we have no values for the 350 items on which the PCA was based.

Discussion

As expected, we found that emotion word properties predict LD performance, over and above lexico-semantic properties. In particular, the results showed that valenced words have a processing advantage over neutral words as predicted. This finding supports the claim that emotional stimuli are processed independently of valence (positive or negative) as in other studies using tasks that do not require an explicit emotional evaluation (e.g., Kanske & Kotz, 2007; Kousta et al., 2009). Although our results are not compatible with models of emotion word processing that assume a difference for positive and negative words (Pratto & John, 1991), the results do endorse a construct of emotionality effects on performance.

Our results also suggest that valence and arousal affect word recognition in an interactive manner, supporting Robinson et al.'s (2004) model. Our study extends previous work however because both emotion dimensions were manipulated and different arousal levels were compared within all valenced words. The present results are in line with the study by Larsen et al. (2008) based on a larger dataset, showing slower LD performance for negative low-arousal words than for negative high-arousal and positive words, thus strengthening the view that valence is modulated by arousal during LD. However, in contrast to Larsen et al. (2008) and also Kousta et al. (2009), ratings of word properties in the present experiment were selected from a single corpus, characterised by homogeneous instructions, i.e., the same standard procedures for rating affective as well as other lexico-semantic properties were used.

Accuracy results showed a different pattern compared to RTs: recognition of positive words was more accurate than negative and neutral words. Given the high accuracy usually obtained with an LD task, even slight differences among conditions become statistically significant but are not necessarily meaningful. Nevertheless, this

difference is in line with previous studies (e.g., Schacht & Sommer, 2009) and could be accounted for by the proposal that positive concepts are better interconnected in the mental lexicon than negative ones (Ashby, Isen, & Turken, 1999). This bias is only reflected in accuracy measures, probably because RTs to negative stimuli need to be equally fast, due to their threatening nature. Another possible interpretation is that the positive words used were perceived by participants as more relevant to themselves than negative and neutral words, in line with observed rating biases (Citron et al., in press; Lewis et al., 2007).

When effects of familiarity and self-reference were removed statistically, the effects of emotionality in LD performance disappeared. This was possibly due to the fact that the emotionality effect was driven by responses to positive words. A self-relevance effect independent of valence was reported by Wentura, Rothermund and Bak (2000), suggesting that perceived relevance might play a role beyond emotional variables.

Importantly, the valence by arousal interaction was confirmed and further supported by participant and item analyses after removing the effects of familiarity and self-reference. **These variables affect the item source of variability and the fact that the results are confirmed also in the item analysis, after control for these possibly confounding factors, makes our findings more robust. The interaction** suggests that manipulating both dimensions provides a more precise and robust window on word processing. In addition, our stimuli were not extreme in arousal: we avoided very highly arousing negative words (e.g., *war*, *rape*) and taboo words because they are difficult to match for arousal with positive words, indeed negative words tend to be higher in arousal (Citron et al., in press). Therefore, effects of emotionality may be limited to very intense stimuli (although the interaction held after additional control).

The post-hoc analyses showed no obvious difference in the effects of emotional variables on word recognition across genders and no interaction of gender with emotional variables (in line with Bauer & Altarriba, 2008). Rather, the pattern of results mirrors the ones found in the original analyses, lending further support to our crucial interaction between valence and arousal dimensions.

We identified effects of affective variables on written word recognition by controlling a wide range of lexico-semantic variables. Nevertheless, since affective and lexico-semantic variables correlate in complex ways, it is difficult to estimate the stability of our results. Replication of this study with different sets of words and a different participant sample, ideally in a different language, is therefore needed. In order to strengthen our results, the same pattern should be obtained through orthogonal manipulation of valence and arousal, i.e. high and low-arousal words should not differ in absolute valence, which was the case in the present study. Further, the gender imbalance in the SAWL norms (Citron et al., in press), as well as in the ratings collected for self-reference, are not ideal in terms of their generalisation to other populations. Finally, the results need to be replicated with words belonging to the same grammatical category, and not to different categories as in our study, since grammatical class is known to affect processing of emotion words (Schacht & Sommer, 2009).

Future work can address the time course of the arousal by valence interaction by means of neurophysiological measures such as event-related potentials (ERPs) and provide further testing of Robinson et al. (2004)'s model, which predicts early implicit integration of each emotional dimension. Some ERP studies have addressed similar issues: one study found interactive effects of emotional variables in early ERP components assumed to index implicit processing stages (Hofmann et al., 2009),

whereas another study found an effect of arousal using only negative words on later components, indexing more controlled and explicit processing stages (Bayer, Sommer, & Schacht, 2010); a third study orthogonally manipulated valence and arousal and reported distinct electrophysiological effects of these variables, but no interaction (Bayer et al., 2012), possibly because their stimuli were all relatively low in arousal.

Conclusions

Our study showed a clear distinction between affective and lexico-semantic variables and an independent contribution from affective variables to predicting word recognition. Emotion words are processed faster than neutral words and an advantage in accuracy for positive over negative and neutral words is reported. Valence and arousal dimensions affect word recognition interactively, in line with an implicit approach-withdrawal emotion processing framework proposed by Robinson et al. (2004): Positive, high-arousal and negative low-arousal words elicit conflicting approach-withdrawal tendencies and so are processed more slowly than positive low-arousal and negative high-arousal words, which elicit congruent approach and withdrawal tendencies, respectively. We contend that both valence and arousal need to be taken into account in studies of emotion word processing because they have an interactive relationship. Furthermore, careful control over correlated lexico-semantic variables must be exerted in future studies.

References

- Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology*, *133*, 323-338. doi: 10.1037/0096-3445.133.3.323
- Ashby, F. G., Isen, A. M., & Turken, U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*, 529-550.
- Bauer, L. M., & Altarriba, J. (2008). An investigation of sex differences in word ratings across concrete, abstract and emotion words. *The Psychological Record*, *58*, 465-474.
- Bayer, M., Sommer, W., & Schacht, A. (2010). Reading emotional words within sentences: The impact of arousal and valence on event-related potentials. *International Journal of Psychophysiology*, *78*, 299-307. doi: 10.1016/j.ijpsycho.2010.09.004
- Bayer, M., Sommer, W., & Schacht, A. (2012). P1 and beyond: Functional separation of multiple emotion effects in word recognition. *Psychophysiology*, *49*, 959-969. doi: 10.1111/j.1469-8986.2012.01381.x
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. [corpus, rating study]. *Behavior Research Methods, Instruments, & Computers*, *33*(1), 73-79.
- Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Simuli, instruction manual and affective ratings. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*, 211-226. doi: 10.1016/j.bandl.2011.12.007
- Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (in press). How are affective word ratings related to lexico-semantic properties? Evidence from the Sussex Affective Word List (SAWL). *Applied Psycholinguistics*. doi: 10.1017/S0142716412000409
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, *12*, 335-359.
- Eckel, C., & Grossman, P. (2002). *Differences in the economic decisions of men and women: experimental evidence*. Amsterdam: Elsevier Science B.V. North-Holland.

- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. Comment reply. *Emotion*, 8, 453-457. doi: 10.1037/a0012887
- Estes, Z., & Verges, M. (2008). Freeze or flee? Negative stimuli elicit selective responding. *Cognition*, 108, 557-565. doi: 10.1016/j.cognition.2008.03.003
- Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74, 967-984.
- Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L.-H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective & Behavioural Neuroscience*, 9, 389-397. doi: 10.3758/9.4.389
- Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, 131, 684-712. doi: 10.1037/0033-2909.131.5.684
- Kahan, T. A., & Hely, C. D. (2008). The role of valence and frequency in the emotional Stroop task. *Psychonomic Bulletin & Review*, 15, 956-960. doi: 10.3758/PBR.15.5.956
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, 1148, 138-148. doi: 10.1016/j.brainres.2007.02.044
- Kemp, A. H., Silberstein, R. B., Armstrong, S. M., & Nathan, P. J. (2004). Gender differences in the cortical electrophysiological processing of visual emotional stimuli. *NeuroImage*, 16, 632-646. doi: 10.1016/j.neuroimage.2003.09.055
- Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112, 473-481. doi: 10.1016/j.cognition.2009.06.007
- Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, 6, 62-72. doi: 10.1037/1528-3542.6.1.62
- Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, 8, 445-452. doi: 10.1037/1528-3542.8.4.445
- Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, 17, 742-748. doi: 10.1093/cercor/bhk024
- McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: The role of fast and slow components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 382-292. doi: 10.1037/0278-7393.30.2.382

- Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*, 136-141. doi: 10.1007/s10339-005-0050-6
- Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: Enhanced sensitivity to negative word valence. *Emotion*, *9*(5), 609-618. doi: 10.1037/a0016350
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. [corpus]. *Journal of Experimental Psychology Monograph Supplement*, *76*, 1-25.
- Pratto, F., & John, O. P. (1991). Automatic vigilance: the attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380-391.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, *55A*(4), 1339-1362. doi: 10.1080/02724980244000099
- Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin*, *30*, 1472-1484. doi: 10.1177/0146167204266647
- Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective & Behavioural Neuroscience*, *9*, 28-43. doi: 10.3758/CABN.9.1.28
- Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*, 1114-1123.
- Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, *80*, 95-104. doi: 10.1016/j.biopsycho.2008.03.010
- Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology*, *78*, 1024-1037. doi: 10.1037/0022-3514.78.6.1024

Table 1.

Descriptive statistics for affective, lexical and semantic features of the stimuli. Mean, minimum and maximum scores for each condition are reported. Emotionality refers to the absolute valence ratings. N-size and N-frequency refer to neighbourhood size and frequency respectively.

	Neutral (50 words)			Positive high arousal (25 words)			Positive low arousal (25 words)			Negative high arousal (25 words)			Negative low arousal (25 words)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Emotionality	.38	.00	.96	1.92	1.01	2.52	1.58	1.06	2.13	1.63	1.16	2.33	1.38	.99	2.07
Emotional Valence	.24	-.73	.96	1.92	1.01	2.52	1.58	1.06	2.13	-1.63	-1.16	-2.33	-1.38	-.99	-2.07
Arousal	2.47	1.50	4.15	4.46	4.04	5.35	3.63	2.84	4.18	4.58	4.02	5.21	3.55	2.61	4.06
Imageability	4.02	2.05	6.57	4.65	2.56	6.38	3.35	2.07	6.33	4.61	3.02	6.18	3.05	1.96	4.61
Age of Acquisition	4.05	1.73	5.23	3.52	1.99	5.63	4.11	1.88	5.54	4.01	2.39	4.90	4.12	2.30	5.21
Log Frequency	1.34	.48	2.44	1.32	.00	2.24	1.24	.30	2.44	1.39	.60	1.98	1.26	.30	2.43
Letters	6	4	10	7	4	9	7	4	10	6	4	10	6	4	10
Phonemes	5	3	10	5	3	8	6	3	9	5	3	8	5	3	9
Syllables	2	1	4	2	1	3	2	1	4	2	1	3	2	1	4
N-Size	3	0	22	3	0	12	2	0	8	3	0	17	4	0	20
N-Frequency	4	0	9	4	0	9	3	0	8	4	0	9	4	0	8

Table 2.

(a) Descriptive statistics of RTs and accuracy rates for Lexicality (words vs. non-words), Emotionality and Valence by Arousal designs (analysis by items). PH = positive high-arousal, PL = positive low-arousal, NH = negative high-arousal, NL = negative low-arousal; (b) Descriptive statistics of familiarity and self-reference ratings broken down by emotionality.

(a)

Design	Condition	mean RT (SE)	mean acc. % (SE)
Lexicality	Words	585.09 (3.24)	98 (.2)
	Non-words	659.46 (4.91)	96 (.5)
Emotionality	Positive	578.60 (5.64)	99 (.4)
	Negative	582.55 (5.19)	97 (.4)
	Neutral	594.11 (5.86)	97 (.4)
Valence by arousal	PH	572.07 (7.39)	99 (.6)
	PL	585.13 (7.39)	98 (.6)
	NH	567.77 (7.39)	98 (.6)
	NL	597.33 (7.39)	97 (.6)

(b)

Condition	Mean	Familiarity		Self-reference		
		Min	Max	Mean	Min	Max
Positive	5,07	3,34	6,62	4,12	2,46	5,84
Negative	4,52	3,43	6,55	2,51	1,49	4,32
Neutral	4,35	2,29	6,51	2,93	1,69	5,88

Table 3.

(a) Factor score coefficients; (b) Rotated component matrix; (c) Regression of factor scores for the 150 words employed in the experiment on the mean LD latency for each word.

(a)					(b)				
Variables	Components				Variables	Components			
	1	2	3	4		1	2	3	4
Letters	.22	.05	-.02	.10	Letters	.81			
Phonemes	.21	.09	-.02	.16	Phonemes	.79			.44
Syllables	.22	.08	-.03	.10	Syllables	.78			
N-Size	-.33	.03	-.01	.28	N-Size	-.79			
N-Frequency	-.38	-.07	.03	.32	N-Frequency	-.81			
Familiarity	.05	.57	.06	.25	Familiarity		.91		
Log Freq	.08	.43	-.07	.02	Log Freq		.77		
AoA	-.08	-.28	.01	.26	AoA		-.66		.51
Emotionality	-.03	.04	.53	.04	Emotionality			.93	
Arousal	-.04	-.06	.53	-.04	Arousal			.93	
Imageability	.19	-.11	.00	-.76	Imageability				-.89

(c)	b	SE b	β	% Variance
(Constant)	583.77	2.32		
Factor 1: LEXICAL (length, neighbourhood)	25.40	2.81	.51***	24.2***
Factor 2: LEXICAL (frequency measures)	-20.59	2.50	-.49***	24.1***
Factor 3: SEMANTIC	-6.76	2.39	-.16**	2.5**
Factor 4: EMOTIONAL	-6.50	2.27	-.17**	2.6**

*** $p < .001$, ** $p < .01$, * $p < .05$

Figure caption

Figure 1. Lexical decision times as a function of Valence and Arousal. (a) Here are shown the estimated marginal means and SEs of the RTs, corrected for familiarity, self-reference and imageability (analysis by item): (a) for the full participant sample; (b) for women only; (c) for men only.