Emotion and language: Valence and arousal affect word recognition

Victor Kuperman[#],
McMaster University, Canada

Zachary Estes[#],
Bocconi University, Italy

Marc Brysbaert, and
Ghent University, Belgium

Amy Beth Warriner
McMaster University, Canada

[#] These authors contributed equally to this work.

Abstract

Emotion influences most aspects of cognition and behavior, but emotional factors are conspicuously absent from current models of word recognition. The influence of emotion on word recognition has mostly been reported in prior studies on the automatic vigilance for negative stimuli, but the precise nature of this relationship is unclear. Various models of automatic vigilance have claimed that the effect of valence on response times is categorical, an inverted-U, or interactive with arousal. The present study used a sample of 12,658 words, and included many lexical and semantic control factors, to determine the precise nature of the effects of arousal and valence on word recognition. Converging empirical patterns observed in word-level and trial-level data from lexical decision and naming indicate that valence and arousal exert independent monotonic effects: Negative words are recognized more slowly than positive words, and arousing words are recognized more slowly than calming words. Valence explained about 2% of the variance in word recognition latencies, whereas the effect of arousal was smaller. Valence and arousal do not interact, but both interact with word frequency, such that valence and arousal exert larger effects among low-frequency words than among high-frequency words. These results necessitate a new model of affective word processing whereby the degree of negativity monotonically and independently predicts the speed of responding. This research also demonstrates that incorporating emotional factors, especially valence, improves the performance of models of word recognition.

Keywords

arousal; automatic vigilance; emotion; lexical decision and naming; valence; word recognition

Corresponding authors: Victor Kuperman, Department of Linguistics and Languages, McMaster University, Togo Salmon Hall 626, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M2, Phone: 905-525-9140, x. 20384, vickup@mcmaster.ca Zachary Estes, Department of Marketing, Bocconi University, Phone: +39-0258-36-36-34, estes@unibocconi.it.

Emotion influences most aspects of cognition and behavior, from visual attention (Rowe, Hirsh, & Anderson, 2007) to social comparison (Estes, Jones, & Golonka, 2012). It affects how we see the world, what we think, and with whom we associate (Forgas, 1995; van Kleef, 2009). Emotions are typically characterized along two primary dimensions of arousal and valence (Russell, 2003; Russell & Barrett, 1999), which correspond respectively to Osgood and colleagues' (Osgood, Suci, & Tannenbaum, 1957) semantic factors of *activity* and *evaluation*. *Arousal* is the extent to which a stimulus is calming or exciting, whereas *valence* is the extent to which a stimulus is negative or positive. These two dimensions are theoretically orthogonal: Negative stimuli can be either calming (e.g., *dirt*) or exciting (e.g., *snake*), and positive stimuli can also be calming (e.g., *sleep*) or exciting (e.g., *sex*). Arousal and valence are also neurologically dissociable, activating distinct cortical networks (Kensinger & Corkin, 2004; LaBar & Cabeza, 2006).

The present research investigates effects of arousal and valence on word recognition. Word recognition has received considerable research attention over the last few decades, and despite a number of important theoretical advances (see Adelman, 2012), a great deal of the variance in word recognition times still remains unexplained (Adelman, Marquis, Sabatos-DeVito, & Estes, 2013). Notably, the current models incorporate a broad range of lexical factors such as word frequency (Brysbaert & New, 2009) and contextual diversity (Adelman, Brown, & Quesada, 2006), but emotional factors are conspicuously absent. So given the influence of emotion on cognition, and the lack of emotional factors in current models of word recognition, the present study examined the influence of emotion on word recognition.

Effects of emotion on word recognition

Many experiments over decades of research suggested that negative stimuli elicit slower responses than neutral stimuli on a range of cognitive tasks. For instance, negative words such as *coffin* tend to evoke slower color naming in the emotional Stroop task (for a review, see Williams, Mathews, & MacLeod, 1996), slower lexical decisions (e.g., Wentura, Rothermund, & Bak, 2000), and slower word naming (a.k.a., reading aloud; e.g., Algom, Chajut, & Lev, 2004) than neutral words such as *cotton*. This observation was attributed to a process of *automatic vigilance*, whereby humans preferentially attend to negative stimuli (Erdelyi, 1974; Pratto & John, 1991). According to this automatic vigilance hypothesis, negative stimuli engage attention longer than other stimuli (Fox, Russo, Bowles, & Dutton, 2001; Ohman & Mineka, 2001), and hence negative stimuli elicit slower responses than other stimuli. The automatic vigilance hypothesis thus assumes that emotion affects the decisional or response stage of word processing: The delayed response to negative words arises during the lexical decision or naming process, rather than during the activation of lexical or semantic representations. Alternatively, emotion could affect the activation of those lexico-semantic representations (Yap & Seow, 2013). That is, activation of negative representations may be "repressed" (Erdelyi, 1974) and/or positive representations may be activated particularly quickly. In fact, Yap and Seow recently reported evidence that valence affects both early and late stages of the word recognition process.

Those decades of experimental results, however, are critically undermined by a lack of stimulus controls (Larsen, Mercer, & Balota, 2006). Larsen et al. conducted a meta-analysis of 1033 stimulus words that were used in 32 published studies on the emotional Stroop task (i.e., color naming of emotional and neutral words). They found that the negative words used in those prior studies tended to be longer and less frequent than the neutral words (see also Warriner, Kuperman & Brysbaert, 2013). These lexical confounds, both of which are known to slow down word recognition (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), could parsimoniously explain the effects observed in those prior studies. And indeed, Larsen et al. found that after controlling those spurious lexical confounds, negative words no longer elicited slower responses than neutral words. Thus, the entire literature on automatic vigilance was rendered equivocal. Since Larsen et al.'s (2006) critical observation, several more recent and better controlled studies have examined the effect of emotion on word recognition, but unfortunately those studies have yielded differing conclusions.

Recent Controlled Studies

Estes and Adelman (2008a) examined the influence of valence on lexical decision and naming latencies, while controlling other important emotional and lexical factors (see Table 1). They found that arousal significantly predicted word recognition: Exciting words tended to be recognized faster than calming words. Their analyses additionally showed that, even after statistically accounting for arousal and several lexical factors, valence still explained significant variance in lexical decision and naming times. Negative words tend to be recognized more slowly than positive words. In contrast to a linear effect whereby increasingly negative and increasingly positive words elicit increasingly slow and fast response times (RTs) respectively, Estes and Adelman found that the effect of valence on word recognition times was nonlinear. Extremely negative words were recognized no slower than moderately negative words, and extremely positive words were recognized no faster than moderately positive words. This produced a step-function whereby RTs remained constant and slow across the category of negative words, decreased sharply through the neutral region of the valence scale, and then remained constant and fast across the category of positive words.

Whereas Estes and Adelman (2008a) tested for independent effects of valence and arousal, Larsen, Mercer, Balota, and Strube (2008) examined whether arousal and valence have an interactive effect on word recognition. They replicated Estes and Adelman's analyses of lexical decision and naming times (except with different control factors, see Table 1), and additionally included the possible interaction between arousal and valence. Larsen et al. found a significant interaction between arousal and valence in lexical decisions (but not in naming), such that low arousal tends to slow down lexical decisions to negative words but speeds up lexical decisions to positive words (see also Robinson, Storbeck, Meier, & Kirkeby, 2004). Highly arousing words, in contrast, exhibited little or no effect of valence. Estes and Adelman (2008b) subsequently demonstrated, however, that Larsen et al.'s reported interaction of valence and arousal depended critically on the underlying form assumed for valence. When valence was entered into the regression model as a linear continuous predictor, then it interacted with arousal in predicting RTs (as in Larsen et al., 2008). However, when valence was entered into the model as a categorical predictor (as

previously observed by Estes and Adelman, 2008a), the interaction reported by Larsen et al. disappeared and negative words elicited slower lexical decisions than positive words regardless of their arousal (i.e., an effect of valence was also observed among highly arousing words).

A limitation of the studies by Estes and Adelman (2008a) and Larsen et al. (2008) was their use of the Affective Norms for English Words (ANEW; Bradley & Lang, 1999) as the sole source of stimuli. ANEW is useful for sampling a limited number of emotional words, but because the words in ANEW were primarily selected for their emotionality, ANEW lacks the preponderance of emotionally neutral words that is typical of natural languages (Kousta, Vinson, & Vigliocco, 2009). Kousta et al. thus merged ANEW with an additional set of randomly selected words, producing a total of 1446 words, including more neutral words than the prior studies. They also employed more sophisticated regression methods for detecting nonlinear relationships. Unlike Estes and Adelman, Kousta et al. found no effect of arousal on lexical decision latencies when controlling for valence. Critically, they also found that after controlling for several other lexical, semantic, and emotional factors (see Table 1), negative and positive words *both* elicited faster lexical decisions than neutral words, and the difference between negative and positive words was nonsignificant. That is, Kousta et al. found a nonlinear, inverted-U effect of valence on lexical decision times. They did not test for an interaction between arousal and valence. These findings based on the large-scale behavioral data set collected in US universities and available from the English Lexicon Project (*ELP*; Balota et al., 2007) have recently been replicated by Vinson, Ponari, and Vigliocco (2013) with the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012), a mega-study that reported lexical decision latencies to over 28,000 words collected at UK universities. Vinson et al. (2013) observed an inverted-U effect of valence on lexical decision times, and they found no evidence of a valence × arousal interaction.

Emotion × Frequency Interactions

Word frequency is among the most important factors of word recognition. To begin with, in most studies it explains a relatively large amount of the variance in word recognition latencies and accuracies (Balota et al., 2004; Brysbaert & New, 2009; Yap & Balota, 2009): Frequent words are recognized more quickly and accurately than infrequent words. More critically for the present study, frequency also tends to modulate the effects of other factors on word recognition. For instance, although both imageability and age of acquisition influence word recognition (Balota et al., 2004; Brysbaert & Cortese, 2011; Cortese & Khanna, 2007; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), both of those effects are significantly larger among low frequency words than among high frequency words (e.g., Cortese & Schock, 2013; Gerhand & Barry, 1999a, 1999b). Two plausible explanations of such interactions with frequency can be differentiated. One general explanation is purely statistical and relies on a base-rate effect, namely, that the magnitude of word recognition latencies is positively correlated with the magnitude of lexical effects on the speed of word recognition. The same relative effect size (say, a 25% difference in RTs between words of high and low imageability) leads to a larger absolute effect in words with longer mean latencies (e.g., 150 ms in words with a mean RT of 600 ms) than in words with shorter mean latencies (e.g., 100 ms in words with a mean RT of 400 ms). Since lower-frequency words

take longer to recognize, all lexical effects may appear larger in those words (Butler & Hains, 1979; Faust, Balota, Spieler, & Ferraro, 1999; Kuperman & Van Dyke, 2013; Yap, Balota, Sibley, & Ratcliff, 2012). A second explanation is that because low frequency words take longer to recognize, there is more time for higher-level semantic factors (e.g., imageability) to affect responding. In contrast, because high frequency words are recognized relatively quickly, semantic factors exert little or no effect on word recognition (Cortese & Schock, 2013). Thus, the former explanation is purely mathematical, whereas the latter is cognitive.

Word frequency also appears to modulate emotional effects on word recognition, but the nature of this modulation is currently unclear. In the emotional Stroop task, valence influenced responses to low frequency words, such that negative words elicited slower color naming than positive words. Among high frequency words, however, valence had no effect (Kahan & Hely, 2008). This finding is analogous to the results described above, in that high frequency tends to reduce or eliminate effects of other factors (e.g., imageability, age of acquisition, valence). In lexical decisions, however, some evidence suggests an opposite effect. Scott, O'Donnell, Leuthold, and Sereno (2009) reported that among low frequency words, negative and positive words elicited equally slow responses, but that among high frequency words, negative words elicited slower lexical decisions than positive words. Further research with eye movements during sentence reading confirmed the interaction of valence and frequency. Whereas fixation durations did not differ between low frequency words of negative and positive valence, fixations on high frequency words were significantly longer for negative words than for positive words (Scott, O'Donnell, & Sereno, 2012). Furthermore, Sheikh and Titone (in press) observed speed benefits to both positive and negative words, as compared to neutral ones, but only when words were of low frequency and relatively concrete. Thus, despite empirical ambiguity in the direction of the effect, it is now clear that word frequency often modulates emotional effects on word recognition. Unfortunately, none of the recent controlled studies of emotional effects on word recognition (i.e., Estes & Adelman, 2008a, 2008b; Kousta et al., 2009; Larsen et al., 2008) controlled or tested for interactions with word frequency.

The Present Study

Empirical Contribution

Prior studies have demonstrated that emotion influences word recognition, but the precise nature of this relationship is unclear. Two main theoretical issues have arisen. First, there is disagreement about the functional form of the effect of valence on word recognition. Specifically, it is unclear whether the effect of valence on word recognition is monotonic but has a step-function form (Estes & Adelman, 2008a, 2008b), monotonic with a linear form (Larsen et al., 2008) or nonmonotonic with an inverted-U form (Kousta et al., 2009; Vinson et al., 2013). Second, it is unclear whether arousal and valence have independent effects on word recognition. Some researchers have found that both arousal and valence influence word recognition (Estes & Adelman, 2008a), whereas others have found effects of valence but not arousal (Kousta et al., 2009). Moreover, some have found that arousal and valence have an interactive effect on word recognition (Larsen et al., 2008), but others have argued

against the validity of such an interaction (Estes & Adelman, 2008b; Vinson et al., 2013). Thus, the present study compared statistical models that varied in whether they treated arousal and valence as linear or nonlinear and independent or interactive.

The prior studies have also exhibited some potentially critical empirical limitations. To begin with, although they included substantially larger samples of stimuli than the pre-2006 experiments in this area of research, those regression studies each sampled little more than a thousand words (see Table 1). By the current standards of research on word recognition (e.g., Brysbaert & New, 2009; Yarkoni, Balota, & Yap, 2008; for review see Adelman, 2012), those are small samples. Moreover, although Kousta et al. (2009) and Vinson et al. (2013) added a few hundred neutral words, the prior studies nonetheless contain a paucity of neutral words, thus undermining their representativeness. Furthermore, the various studies have included different sets of control factors (see Table 1), making it difficult to compare results from one study to the next. For instance, the discrepant results between Estes and Adelman (2008a) and Kousta et al. could simply be due to the fact that Estes and Adelman did not control for age-of-acquisition, or that Kousta et al. did not control contextual diversity. Perhaps most importantly, those prior studies did not test for emotion × frequency interactions, which are known to occur in word recognition (Kahan & Hely, 2008; Scott et al., 2009, 2012; Sheikh & Titone, in press). The present study aimed to address these limitations by (1) sampling a substantially larger set of words that (2) were not sampled for their emotionality and thus are more representative of natural language, (3) including many more lexical and semantic control factors than any prior study, and (4) testing for interactions of valence and arousal with word frequency.

Thus, the present study used a sample (12,658 words) from the dataset of affective norms (psychological valence, arousal and dominance) collected by Warriner et al. (2013), which is about 9-13 times larger than the prior studies. The analyses also included about twice as many lexical and semantic control factors, to critically test multiple models of emotional word recognition. Our study has been made possible by the recent emergence of psycholinguistic mega-studies (for review see Adelman, 2012; Balota, Yap, Hutchinson, & Cortese, 2012), whereby massive datasets compiling the lexical (Brysbaert & New, 2009; Kuperman et al., 2012), semantic (Brysbaert, New, & Keuleers, 2012), and emotional characteristics (Warriner et al., 2013) of many thousands of words can be merged with behavioral data such as lexical decision and naming latencies and accuracies for those same words, namely the English Lexicon Project (Balota et al., 2007). The present research employs this mega-study approach to determine the precise nature of the effects of arousal and valence on lexical decision and naming latencies.

Theoretical Contribution

We anticipate two important theoretical contributions from this research. First, this research is most informative for models of automatic vigilance. The effect of valence on word recognition times has been a primary source of evidence for automatic vigilance (Algom et al., 2004; Estes & Adelman, 2008a, 2008b; Larsen et al., 2008; Pratto & John, 1991; Wentura et al., 2000; Williams et al., 1996), but several different relationships have been hypothesized. The simplest model of automatic vigilance supposes that humans immediately

judge stimuli as either aversive (i.e., negative stimuli to be avoided) or appetitive (i.e., positive stimuli to be approached) in a binary, categorical manner (Estes & Adelman, 2008a, 2008b). Such simple evaluative judgments would be behaviorally adaptive in that they would facilitate rapid decisions and actions. Deliberating about whether a stimulus is extremely dangerous or only moderately dangerous could in fact be fatal, whereas over-reacting with an extreme response to a moderately dangerous stimulus is merely disruptive rather than fatal. Of course, humans are capable of differentiating extreme from moderate stimuli, but the implication is that such fine discriminations occur via a slower, more deliberative process than the one that influences word recognition times. By this categorical model of vigilance, the relation between valence and recognition times should be a step function, with slower responses to negative words than to positive words (Estes & Adelman, 2008a, 2008b).

A different model arises if humans do make use of the fine discrimination of negative, neutral and positive valence, such that it differentially affects either how fast a stimulus activates its lexical or semantic representation (slower for negative words) or how long it engages attention (longer for negative words) or both. By this account, a gradient effect of automatic vigilance is expected. The gradient model predicts a linear negative effect of valence on behavioral latencies, with slower responses to more negative words and a speed-up with an increase in valence. Somewhat surprisingly, none of the recent controlled studies supported such a gradient model of automatic vigilance.

In contrast to prior evidence that negativity *slows down* word recognition (Algom et al., 2004; Estes & Adelman, 2008a, 2008b; Pratto & John, 1991; Wentura et al., 2000; Williams et al., 1996), Kousta et al. (2009) found that valence, whether negative or positive, *sped* word recognition. Such an inverted-U relation between valence and recognition times would entail a double rejection of automatic vigilance: Responses are claimed to be (1) faster to negative words than to neutral words, and (2) equally fast to negative and to positive words. Kousta et al. instead explain their result in terms of motivational relevance: Because negative and positive stimuli respectively activate the avoidance and approach behavioral systems, both valences are "motivationally relevant," and motivationally relevant stimuli are preferentially processed (Lang, Bradley, & Cuthbert, 1990).

Alternatively, an interaction of valence and arousal (Larsen et al., 2008) would imply yet a different model of vigilance. In fact, such an interaction effect on word recognition times would corroborate some prior research on evaluative judgments. Robinson et al. (2004) presented images and words that varied in arousal and valence, and had participants indicate whether the stimulus was negative or positive. They found a similar interaction as that observed by Larsen et al.: Negative words tended to elicit faster responses when they were highly arousing than when they were calming, whereas positive words elicited faster responses when they were calming than when they were arousing. According to Robinson et al., high arousal facilitates responding to negative stimuli because this combination of arousal and negativity is characteristic of dangerous stimuli, and rapid responding to dangerous stimuli is adaptive. Thus, by examining the precise nature of the effects of arousal and valence on word recognition latencies, the present research provides a critical test of various models of automatic vigilance.

Secondly, this research may also inform models of word recognition. Lexical and semantic factors such as word frequency and age of acquisition have long been known to influence the speed with which words are recognized, and decades of research have identified a substantial list of factors that each explain some significant amount of variance in word recognition times. For instance, Adelman et al. (2013) recently assembled a regression model that included a comprehensive list of such factors, and the regression model outperformed all current cognitive models of reading. In so doing, however, Adelman et al. highlighted how little of the potentially explainable (i.e., non-noise) variance is actually explained by the current knowledge in the field. Essentially, Adelman et al. announced a call for the field to search for additional factors or alternative models that can more fully explain word recognition. One class of likely predictors of word recognition missing from current models is emotional factors, which could influence the early activation of lexico-semantic representations and/or the late decisional-response stage of word processing (Yap & Seow, 2013). Thus, by testing for effects of valence and arousal on word recognition, the present research contributes generally to models of word recognition.

Methods

Data

We compiled a set of 12,658 words for which all of the following variables were available.

Emotion variables—Mean *valence* and *arousal* ratings, retrieved from Warriner et al. (2013), served as our predictor variables of primary interest.

Behavioral variables—Mean *lexical decision and naming latencies*, retrieved from the ELP (Balota et al., 2007), served as our criterion variables.

Lexical control variables—*Word length* was controlled via several measures: Orthographic length in characters and morphemes, and phonological length in phonemes and syllables. *Lexical density* was also controlled via several measures: Orthographic, phonological and phonographic neighborhoods (these are the number of words that can be formed from a given word by replacing respectively one letter, one phoneme, or one letter corresponding to one phoneme, with another in its place), and orthographic and phonological Levenshtein distance (OLD and PLD; these are defined as the mean Levenshtein distance between a target word and its 20 closest neighbors, where Levenshtein distance is the minimum number of letter/phoneme insertions, deletions, or substitutions needed to transform the target word into another word). All these values were retrieved from the ELP. *Word frequencies* were retrieved from the 51 million-token SUBTLEX-US corpus of subtitles to the US films and media (Brysbaert & New, 2009), the 130 million-token HAL corpus of electronic communication (Burgess & Livesay, 1998), and the 8 million-token TASA12 corpus of educational materials for 12th graders (Zeno et al., 1995). *Contextual diversity* was also retrieved from SUBTLEX-US, and *age-of-acquisition* (AoA) was retrieved from the norms of Kuperman et al. (2012). We further included the word's *initial phoneme* and its *part-of-speech* (i.e., dominant PoS tag in Brysbaert, New, & Keuleers, 2012).

Statistical Analyses

As demonstrated by Larsen et al. (2008), arousal and valence may enter into interactions that form complex surfaces in the three-dimensional space with arousal, valence and behavioral latency as axes. Recent reports by Kahan and Hely (2008), Scott et al. (2012) and Sheikh and Titone (in press) additionally suggest the possibility of emotion × frequency interactions. These observations necessitate the use of a statistical technique that enables flexible modeling of complex surfaces, without imposing the planar functional form on interactions. Generalized additive mixed-effects (GAM) regression modeling (see e.g., Hastie & Tibshirani, 1990; Wood, 2006) – as implemented in the mgcv package (Wood, 2006, 2011) of the R statistical computing software (R Core Team, 2012) – affords the required flexibility and hence is the regression technique of choice here.[1]

The distributions of raw lexical decision and naming latencies showed the typical skew (i.e., a heavy right tail), which biases estimates of the mean. A common solution is to transform the distribution such that it closely resembles the Gaussian, and to apply statistical methods that assume an underlying Gaussian distribution of the data (see e.g., Baayen & Milin, 2010; Kliegl et al., 2010). In keeping with this approach, we log-transformed the latencies, as indicated by the Box-Cox transformation test (Box & Cox, 1964). Regression models were thus fitted to log-transformed RTs with Gaussian as the underlying family of distributions and identity as a link function. The results reported below were also obtained with both untransformed RTs and inverse-transformed RTs, so our conclusions are not particular to the transformation itself.

We frame our discussion of the functional form of emotion effects in terms of (non)monotonicity rather than (non)linearity, because a linear effect of a predictor on a log-transformed dependent variable only guarantees a monotonic, not necessarily linear, effect. Moreover, the ratings of valence and arousal are ordinal variables, whereas claims of a linear relationship require variables that are at least interval. Therefore, our research question is better thought of as addressing the question whether the emotion effects on word recognition are monotonic with a (near-)constant rate of change across the entire range, or have a specialized form, such as the step-function, indicating a fast change over a limited part of the continuum and a lesser change in the remainder.[2]

Multicollinearity of predictors in a regression model may inflate standard errors and distort regression coefficients (Mason & Perreault Jr., 1991). In the present set of variables, strong correlations typically exist both within and between measures gauging the rate and time-course of word use (frequency of occurrence, contextual diversity, AoA) and measures gauging formal lexical properties (e.g., length in characters, morphemes, phonemes and syllables; orthographic, phonological and phonographic neighborhood sizes, as well as PLD and OLD). Unsurprisingly then, the condition number test calculated for the entire set of

---

[1]For detailed description and worked examples of the use of GAM models in psycholinguistics see Baayen, Kuperman, and Bertram (2010), Tremblay and Baayen (2010), Matuschek, Kliegl, and Holschneider (2012), Kryuchkova et al. (2012), and Balling and Baayen (2012), and for applications in linguistic studies see Wieling et al. (2011) and Koesling et al. (2012).
[2]We thank Stephen Lupker for this suggestion.

continuous variables under consideration (frequency-related measures, length-related measures, valence and arousal) indicated substantial multicollinearity, $\kappa = 87.13$.

Several steps were taken to reduce multicollinearity. First, we applied principal components (*PC*) analysis to the nine variables representing formal lexical properties. Three principal components each explained over 5% of the variance in those formal lexical variables, and taken together accounted for over 90% of the variance. These principal components (labeled *PC1*, *PC2*, and *PC3*) were thus incorporated into our models as statistical estimators of formal lexical properties. (Variables that loaded most strongly on PC1 were length in characters, and orthographic and phonological density; on PC2 – orthographic, phonological and phonographic neighborhoods; and on PC3 – length in morphemes.) Second, the effect of word frequency (from SUBTLEX, log transformed) was partialled out from AoA and log contextual diversity estimates. The residual values (labeled rAoA and rCD) were thus de-correlated from the estimates of frequency and were used in further modeling. Finally, we centered all numerical predictors. The resulting set of PC1, PC2, PC3, rAoA, rCD, word frequency, valence, and arousal variables showed only a mild, acceptable level of multicollinearity, $\kappa = 16.85$.

The set of continuous predictors listed above, as well as factors reflecting the first phoneme and part-of-speech, were entered into GAM models with by-item average RTs as the dependent variable. All continuous predictors were first explored for nonlinear effects, implemented as restricted cubic splines. Predictors that showed no support for a nonlinear functional form were re-entered into final models as linear. We also modeled interactions (implemented as tensor product splines) for predictors that were shown or hypothesized to interact in prior research (i.e., valence $\times$ arousal, frequency $\times$ valence, and frequency $\times$ arousal). Because the dependent variables were by-item average RTs, there were no random effects in any models fitted to the item-level data.

Results

Our analyses addressed a progressive series of research questions, reported in turn.

What is the functional relation between word frequency and emotional factors?

Various corpora have been used for estimating word frequencies in prior studies. However, they differ in potentially relevant ways (e.g., content and size), and indeed they are not equally good at predicting word processing times (Brysbaert & Cortese, 2011; Brysbaert & New, 2009). We therefore first examine whether the various corpora yield systematically different patterns of word frequency estimates across the ranges of valence and arousal. Figure 1 demonstrates the functional relationship of valence and arousal with word frequency estimates from TASA12, SUBTLEX, and HAL. The figure is based on 12,092 words overlapping between the three corpora. The vertical separation among the lines simply reflects the differing sizes, and hence the differing absolute word frequencies, of the various corpora: TASA12 and HAL respectively are the smallest and largest of the three corpora, so they respectively yield the lowest and highest frequency counts. Valence and arousal are binned into twenty quantiles, each accounting for 5% of the respective distribution, and the mean log frequency is reported for each bin.

Frequency distributions across the valence range (left panel) are similar across corpora. While there is an overall trend for more positive words to be more common within each of the three corpora (i.e., all three lines peak on the right end of the scale), very negative words are more frequent than moderately negative words. The observed spike in frequency of very negative words will become important in our comparison of prior and present findings. The functional relationships of frequency and arousal (right panel) differ substantially. TASA12 contains mostly low-arousal words, with highly arousing words being relatively rare, as indicated by a frequency curve that decreases sharply across the arousal range. Put simply, educational texts (TASA12) contain boring words, possibly due to editorial requirements to what counts as appropriate content for school-level reading. SUBTLEX, in contrast, shows a relatively flat pattern across the arousal range, with an increase in frequency in very arousing words. Film and media subtitles (SUBTLEX) thus unsurprisingly contain more exciting words, as befits their purpose of attracting and maintaining viewers' attention. Finally, HAL exhibits an essentially flat distribution of frequency over the arousal range. That is, electronic communication (HAL) contains an approximately equal number of boring, neutral and exciting words. In what follows we only consider SUBTLEX and HAL frequency estimates, as these two corpora are larger and show a stronger convergence than the TASA frequency counts which are based on a (6 to 16 times) smaller sample of edited educational materials.

What is the relation between emotional factors and word recognition when the emotion × frequency interaction is *not* taken into account?

Several recent studies have indicated that emotion may interact with word frequency in affecting word processing (Kahan & Hely, 2008; Scott et al., 2009, 2012; Sheikh & Titone, in press), but the prior regression studies did not include emotion × frequency interactions. For comparison with those prior regression studies, we thus examined such non-interactive relationships between emotional factors and behavioral latencies in our larger and more representative dataset. We plotted valence and arousal against lexical decision and naming response times: as shown in Figure 2, we replicated the inverted-U effect of valence on response times, as originally shown by Kousta et al. (2009). Importantly, the inverted-U shape of the valence effect was retained after statistically accounting for all of the control variables listed in the Methods (plot not shown). These control variables included word frequency (SUBTLEX) but not its interactions with valence and arousal. Unlike Kousta et al., however, our analysis also revealed an inverted-U effect of arousal on response times. Thus, when the hypothesized interactions of word frequency with valence and with arousal were omitted from the analyses (as in prior studies), the inverted U-shaped relationship between emotional factors and behavioral latencies was replicated.

What is the relation between emotional factors and word recognition when the emotion × frequency interaction *is* taken into account?

Figure 3 summarizes the effects of valence (top row) and arousal (bottom row) on lexical decision (left column) and naming (right column) response times, plotted as a function of word frequency (SUBTLEX). Each panel displays a series of five trend lines estimated using the cubic spline function for words falling into respective quintiles of lexical frequency (from a solid line for the lowest frequency words to a dotted line for the highest

frequency words). The top panels reveal that the effect of valence on behavioral latencies is negative, and the magnitude of the effect is attenuated as frequency increases (i.e., the slope is steep among the high lines but is flat in the lowest line). To illustrate, the magnitude of the effect of valence on lexical decision times (top left panel) was about 55 ms among the lowest frequency words, but among the highest frequency words valence had little or no effect. The bottom panels of Figure 3 reveal that the effect of arousal on behavioral latencies is instead positive, and again the magnitude of the effect is attenuated as frequency increases. In the extreme, the magnitude of the effect of arousal on lexical decision times (bottom left panel) was about 55 ms among the lowest frequency words, but among the highest frequency words arousal had little effect.

The patterns in Figure 3 are based on raw data, and are fully confirmed by the regression model that includes emotion × frequency interactions (see models below, plots not shown). The consistent near-linear trends observed in all frequency bands (Figure 3) reveal that the inverted-U shape (Figure 2), which is only observed when emotion × frequency interactions are unaccounted for, substantially mischaracterizes the effect of emotion on word recognition behavior. Finally, the patterns in Figures 2 and 3 are based on SUBTLEX frequencies, but those same patterns are also observed when HAL frequency counts are used instead (plots not shown). Thus, despite being independent corpora based on different genres of text (i.e., film and media subtitles; internet communications), SUBTLEX and HAL frequencies yielded strikingly similar emotion × frequency interactions. Full results of GAM regression models are reported next, separately for lexical decision and naming.

Lexical Decision—A model fitted to lexical decision RTs identified a number of outliers (1.53% of the data points) that were further than 2.5 standard deviations from the model's fitted values (Baayen & Milin, 2010). These outliers were removed and the model refitted. Table 2 reports the model's outcome. Part A of the table lists the linear effects of continuous predictors. For brevity, the effects of factorial predictors with multiple levels – namely, part-of-speech and first phoneme – were omitted from the table. However, both of these control factors were significant, and the full model's output is available upon request. Part B lists the nonlinear effects (i.e., smooth terms, for which the assumption of nonlinearity was warranted, $p < 0.001$ and effective degrees of freedom $edf > 1$) and the emotion × frequency interactions (i.e., tensor products)[3]. The model explained 60.15% of the variance in latencies.

F-test model comparisons were conducted to establish whether the valence × frequency tensor product, and separately the arousal × frequency tensor product, significantly improved the performance of the baseline model with nonlinear non-interacting effects of frequency, valence, and arousal. Both tensor products were indeed warranted as terms in the best-performing model (Table 2), with all $ps < 0.001$ in model comparison tests. Furthermore, the tensor product of frequency and valence was preferred by the model comparison test over the independent nonlinear effect of valence (which was significantly

---

[3]The output of the generalized additive models differs from outputs of most regression or ANOVA models in that the estimates and inferential statistics for tensor products are reported for the entire hyperbolic surface, without separating it into more customary separate representations of main effects and interactions. The main effect of frequency is not omitted, but rather is fully accounted for when frequency is entered as one of terms in the tensor product with valence, arousal or any other variable.

negative, $p < 0.001$). Likewise, the tensor product of frequency and arousal was preferred over the nonlinear effect of arousal (which was nonsignificant, $p = 0.15$). In short, adding the frequency × valence and frequency × arousal interactions significantly improved the fit of the models. The tensor product of valence and arousal did not reach significance in any of the models, suggesting that these affective properties have independent effects.

Critically, including these emotion × frequency interactions revealed effects (Figure 3) that are strikingly different from those observed when the interactions are excluded from the models (Figure 2): Namely, what previously appeared as inverted U-shaped effects of valence and arousal on response times are now revealed to actually be monotonic, essentially linear effects. In none of the frequency bands did the effect of valence on response times exhibit an inverted U-shape. The effect of arousal on response times was also monotonic and near-linear, rather than inverted U-shaped.

There was no straightforward way to estimate the unique variance explained by either valence or arousal, as their impact was modulated by frequency. As an approximate estimate, we compared the amounts of variance explained by (a) the nonlinear effect of frequency, (b) the tensor product of frequency and valence, and (c) the tensor product of frequency and arousal. Models with predictors outlined in (a)-(c) were fitted to RTs from which effects of all other predictors (principal components PC1, PC2, and PC3, AoA, contextual diversity, first phoneme, and dominant part-of-speech) were partialled out. Frequency alone (a) explained 24.4% of the variance, including the frequency × valence interaction (b) explained 26.3%, and including the frequency × arousal interaction (c) explained 24.5% (including both interactions together explained 26.4%). We conclude that the contribution of valence to explained variance (the difference between (a) and (b)) is on the order of 2%, while the contribution of arousal (the difference between (a) and (c)) is much smaller (0.1%).

Naming—The modeling procedure was repeated with naming latencies. Table 3 reports the model fitted to log-transformed (base e) naming latencies after removing outliers (1.82% of the data points). Part A of the table again lists the linear effects of continuous predictors, whereas Part B lists the nonlinear effects and the emotion × frequency interactions. Again, for brevity, the effects of part-of-speech (nonsignificant) and first phoneme (significant) were omitted from Table 3, but the full model's output is available upon request. The model explained 58.01% of the variance in latencies. As with lexical decisions, F-test model comparisons indicated that both tensor products (frequency × valence and frequency × arousal) significantly improve the model's performance as compared to a set of non-interacting, nonlinear effects of frequency, valence and arousal (all $ps < 0.01$). Once again, the interaction of valence and arousal was nonsignificant ($p = 0.3$), pointing to the independent nature of these effects.

Amounts of variance in naming latencies explained by valence and arousal, with all other effects partialled out, were as follows. Frequency alone (a) explained 11.1% of the variance, including the frequency × valence interaction (b) explained 11.3%, and including the frequency × arousal interaction (c) explained 11.2% (including both interactions together explained 11.5%). Thus, in naming the contribution of valence to explained variance is a

small but significant 0.2%, while the contribution of arousal is an even smaller but still significant 0.1%.

Are emotion effects robust across individual trials?

The preceding analyses, and indeed all prior studies, examined emotion effects at the level of words (or "item means"): Each word has a mean response time, and among the set of words, we test whether the words' valence and arousal ratings tend to predict their mean response times. This analysis provides a general, averaged view of emotion effects on word recognition. Here we additionally examine emotion effects at the level of individual trials: Each trial of the lexical decision and naming studies produces a single response latency, and among all those individual trials, we test whether the given word's valence and arousal ratings tend to predict the individual response times that the given word elicited by each participant. To illustrate, suppose a hundred words are presented to each of a hundred participants in a lexical decision study. In the standard word-level analysis (a.k.a. "item analysis", "by-items analysis", or "F2"), there would be 100 rows of data (one per item). But in the trial-level analysis, there is a row for each trial of each participant, so there would be 10,000 rows of data (100 x 100). Clearly, this trial-level analysis is far more statistically powerful, though it must be noted that individual response latencies are also far more variable (due to random "noise" that is averaged out of word-level analyses).

As in previous analyses, only correct responses were considered, and we excluded outliers identified in the ELP data (Balota et al., 2007) as trials with latencies more than 3 standard deviations from the word's mean latency. The resulting data sets contained 384,113 and 329,871 data points for lexical decision and naming respectively. Our models (not shown) had the same configuration of predictors as outlined above, with an addition of such predictors as the latency and correctness of the previous response, and the position of the word in the participant's experimental list. The maximal random effects structure was implemented in the models, with by-subject and by-word intercepts, as well as by-subject slopes for valence, arousal, and frequency and their interactions (Barr et al., 2013). The results were very similar to the ones observed in average latencies. Namely, both lexical decision and naming latencies monotonically decreased with increasing valence, while the valence effect was at its strongest in the lower-frequency words and gradually diminished in magnitude as word frequency increased. The same attenuation of effect with increasing frequency was observed for the positive correlation of arousal with lexical decision and naming latencies. Finally, the valence × arousal interaction did not reach significance in the trial-level data, over and above the frequency × valence and frequency × arousal interactions. Thus, the trial-level analysis replicated the emotion effects observed in the word-level analysis reported above.

Are emotion effects on word recognition independent of semantic variables?

Our preceding analyses included a large number of lexical control factors, but recently several measures of additional semantic factors have emerged. Most pertinently, there is growing interest in "semantic richness", which is essentially the amount or diversity of information that a given word evokes. For instance, "dog" tends to evoke a rich array of sensory and encyclopedic information, whereas "twig" tends to evoke less information. It

could reasonably be argued that emotion is merely one facet of semantic richness, and thus the question arises whether emotion effects on word recognition are really just another demonstration of semantic richness effects. We therefore examined the correlations of valence and arousal with a battery of semantic measures, and we tested whether these emotional factors explained any unique variance in word recognition times after statistically accounting for those semantic variables. For this analysis we identified a set of 1083 monosyllabic words for which all of the following measures were available: valence and arousal ratings (Warriner et al., 2013), SUBTLEX frequency of occurrence (Brysbaert & New, 2009), age-of-acquisition ratings (Kuperman et al., 2012), imageability ratings (Cortese & Fugett, 2004; Schock et al, 2012), sensory experience ratings (Juhasz, Yap, Dicke, Taylor, & Gullick, 2011; Juhasz & Yap, 2013), body-object interaction ratings (Tillotson, Siakaluk, & Pexman, 2008), semantic diversity measures (Hoffman, Ralph, & Rogers, 2012; see also Jones, Johns, & Recchia, 2012) and the word's number of senses from Wordnet (Miller, 1995).

Table 4 demonstrates that although all correlations were weak in magnitude ($|\rho| < 0.2$), all were significant ($p < .05$) except those of arousal with semantic diversity and with the number of senses. Based on these correlations as well as ones reported in Warriner et al. (2013, Table 5) we observe that positive words are consistently associated with higher semantic richness: they are more concrete, imageable, sensorily acute, prone to be used in body-object interactions, etc. To evaluate the amount of variance explained by each of these affective and semantic variables, we calculated the difference in multiple R^2 between a model with non-linear functions of word length, log frequency and age-of-acquisition and a model which included those same predictors plus a non-linear function of one of the variables under comparison. All models were fitted to log-transformed lexical decision latencies. Inclusion of arousal explained an extra 1.4% of the variance (54.4% vs 53%), and valence explained an extra 1.1%. These increments were significant ($p < 0.01$) and stronger than those associated with most other semantic variables: Sensory experience ratings 0.3%, semantic diversity 0.2%, number of senses 0.1%, imageability 0.6%. The amount of variance explained by body-object interactions (1.1%) was on par with that of valence, and smaller than that of arousal. Finally, we observed a significant increment of R^2 when valence was added to form a tensor product with word frequency in the model that additionally had as predictors nonlinear functions of word length, AoA, and all semantic variables listed above. The amount of unique variance associated with valence, calculated over and above the influence of all semantic predictors, was 1.1% (56.1% vs 55%). The comparable quantity for arousal was 1.2%.

We conclude that the independent impacts of valence and arousal cannot be ascribed to their correlations with a large range of semantic variables (these correlations were weak). Nor can those emotion effects be attributed to the variance the affective measures share with the semantic richness measures: The contributions of both valence and arousal are independent of and stronger than those of the semantic variables, and are numerically the same regardless of whether they are estimated over and above the other semantic variables.[4]

Discussion

Converging empirical patterns observed in word-level and trial-level data from lexical decision and naming RTs in American English yield the following conclusions.

1. Valence has a monotonic effect on word response times, such that negative words (e.g., *coffin*) tend to be responded to more slowly than neutral words (e.g., *cotton*), which tend to be responded to more slowly than positive words (e.g., *kitten*). Specifically, the underlying functional form of this relation between valence ratings and log-transformed RTs was strictly linear in the regression analyses we ran; using a curvilinear form for valence failed to improve the fit of the model to the data. Note however that, because the precise statistical properties of the valence scale are currently unknown and because the RTs were log transformed, the linear nature of this effect must be interpreted with caution. What can be concluded with more confidence is that the effect is monotonic and thus constant in polarity across the entire range: Greater negativity generally slows lexical decision and naming RTs.

2. Arousal has a monotonic effect on word response time, such that calming words (e.g., *sleep*) tend to be responded to more quickly than arousing words (e.g., *sex*). That is, arousal slows word processing. As with valence, the relation between arousal and RT was strictly linear in our analyses, but again due to potential nonlinearities in the valence scale and/or the log-transformed RTs, we conclude only that the effect is monotonic.

3. Valence has a stronger effect on word processing than does arousal. Valence explains about 2% of the variance in lexical decision times and 0.2% in naming times, whereas the effect of arousal in both tasks is limited to 0.1% in the analysis of the full dataset.

4. The effects of valence and arousal on word response times are independent, not interactive. Adding an arousal × valence interaction term to the model failed to improve its fit, even when the interaction was flexibly modeled as a hyperbolic surface.

5. Valence and arousal both interact with word frequency, such that valence and arousal exert larger effects among low-frequency words than among high-frequency words.

6. Valence and arousal have stronger effects on lexical decisions than on naming. Valence and arousal together explained more than 2% of the variance in lexical decision latencies, whereas their effects on naming latencies were less than .5%.

Empirical Integration

Our results support many prior findings. Specifically, results 3, 5, and 6 corroborated prior studies showing respectively that valence is more powerful than arousal (see Table 1; see

---

[4]A slightly more prominent predictive role of arousal, as compared to valence, in the subset of 1083 monosyllabic words is intriguing given arousal's negligible role in the entire data set of over 12,000 mono- and polysyllabic words. We link this inflation in the predictivity of arousal in the smaller dataset to the fact that monosyllabic words, as compared to the full word set, are significantly shorter in length (4.36 vs 7.21 characters), higher in (log 10) frequency (2.70 vs 1.99), higher in valence (5.16 vs 5.08) and lower in arousal (4.05 vs 4.20), among other differences (all $ps < 0.01$). The discrepancy serves as another argument against selecting data samples that differ in relevant ways from the language's lexicon as found "in the wild". In this case, a consideration of an exclusively or even predominantly monosyllabic data set would lead to a perception of arousal as a stronger predictor than it proves to be in a more exhaustive analysis.

also Adelman & Estes, 2013), that both interact with frequency (Kahan & Hely, 2008; Scott et al., 2009, 2012; Sheikh & Titone, in press), and that they affect lexical decisions more than naming (Estes & Adelman, 2008a; Larsen et al., 2008). On the other hand, our findings 1, 2, and 4 are novel and inconsistent with some prior results. We consider each of these empirical discrepancies in turn.

The observed functional form of the valence effect is novel and contradicts prior claims that this effect is either a step function or an inverted-U function (Estes & Adelman, 2008a; Kousta et al., 2009; Vinson et al., 2013). Our additional analyses indicate that this discrepancy is likely due to a combination of factors. First, the present dataset is much (9-13 times) larger than the ones used in previous studies. This advantage yields a more natural representation of the ranges of frequency, arousal and valence; a more precise account of nonlinear functional relations between frequency, valence and arousal; and a higher accuracy of estimated curves and hyperbolic surfaces that characterize the effects of emotional variables over and above frequency and other statistical controls. One aspect that a larger dataset may have remedied is an over-representation of extremely negative words in prior studies (Estes & Adelman, 2008a; Kousta et al., 2009; Vinson et al., 2013). Those studies were based on an original or slightly extended ANEW data set, which was specifically developed to include a preponderance of emotional words. To illustrate, whereas the extremely negative words (i.e., those with a mean rating of less than 2 on a 1-to-9 scale) constitute 4.8% of the ANEW sample, they constitute only 0.7% of the Warriner et al. (2013) sample. That is, the relative frequency of extremely negative words is about 7 times higher in ANEW than in Warriner et al.'s randomly sampled word set that we use here. Yet very negative words come with a spike in frequency in all of the three corpora considered (Figure 1): for instance, the bottom 5% bin of the valence distribution (valence: 1.34-2.76) has a higher mean log frequency than any single bin between 5 and 35% of the valence distribution (valence: 2.77-4.74). The over-representation of relatively frequent words in the narrow very negative subrange of valence may have led to the attribution of the response speedup in negative words to the valence effect, whereas it is in fact due to the effect of frequency.

Second, ours is the first study to consider interactions of frequency and emotion in lexical decision and naming. We show that the inverted-U shape of the valence and arousal effects is only observed when emotion × frequency interactions are not accounted for in the analysis (Figure 2). When considered in specific frequency bands, valence and arousal show monotonic near-linear effects, and never the inverted U-shaped effects (see Figure 3). The same monotonic effects are also observed when word frequencies are estimated from HAL instead of SUBTLEX. This suggests, again, that the inverted-U shape may be an artifact of skewed distributions of frequency across the valence range, with higher frequency associated both with very negative and very positive words. The interactions in which strong effects of emotion are observed in low-frequency bands (negative for valence, and positive for arousal) and attenuating effects are observed in words of increasing frequency dovetails perfectly with earlier findings that effects of imageability, age-of-acquisition and other lexical variables are the strongest in lowest-frequency words (e.g., Cortese & Schock, 2013; Gerhand & Barry, 1999a, 1999b).

The monotonic positive effect of arousal is also novel: Kousta et al. (2009) found no effect of arousal, and although Estes and Adelman (2008a) did obtain significant effects of arousal, those effects were in the opposite direction to the effect observed here. The fact that Kousta et al. (2009) found no effect of arousal is unsurprising, considering the extremely small magnitude of the effect that we observed here. The fact that Estes and Adelman (2008a) found a negative effect of arousal can be explained by differences in corpora used to estimate word frequencies. Figure 1 shows that highly arousing words are relatively more frequent in the SUBTLEX and HAL corpora than the TASA12 corpus (i.e., films and websites are more exciting than textbooks). This underestimation of the frequency of high arousal words in TASA12 as compared to SUBTLEX or HAL corpora leads the statistical models to misattribute the facilitative effect of their frequency to a facilitative effect of arousal instead. Because the frequency underestimation is at the high end of the arousal range, this produces an erroneously negative effect of arousal on word recognition. However, when the relatively high frequency of high arousal words is fully accounted for (via SUBTLEX or HAL frequencies), the relation between arousal and word recognition is shown to be positive rather than negative (see Figure 3).

Finally, the independent nature of the valence and arousal effects is novel and fails to replicate the interaction reported by Larsen et al. (2008) in lexical decisions, though it is in line with Vinson et al.'s (2013) findings. While we cannot identify the exact source of discrepancy, it is may stem from our more accurate estimation of effects and interactions due to a larger dataset, the use of hyperbolic surfaces rather than planes in the three-dimensional space to approximate interactive terms, and finally, from our consideration of emotion × frequency interactions, which could have absorbed the variance otherwise attributable to valence × arousal interactions.

One may reasonably wonder, then, why our results should be preferred over prior studies. First, it must be noted that the three preceding studies in Table 1 were *not* independent analyses. Larsen et al. (2008) analyzed the same dataset as Estes and Adelman (2008a), and Kousta et al. (2009) also analyzed a largely overlapping dataset with about 70% of the same stimulus words. So even in cases where our result differs from all three prior studies – as in the arousal effect – this should not be counted as three observations weighed against one observation, because those three observations were based effectively on a single dataset that was analyzed in three ways. Second, whereas the stimuli in prior studies were sampled for their emotionality, the stimuli in the present study represent all words rated as known by at least 70% of raters in the norming study of Kuperman et al. (2012), and without regard for their emotionality. Thus, our sample of stimuli presumably is more representative of natural language. Third, our stimulus sample is about 10 times larger than the previous studies. So again, our stimuli presumably are more representative. Fourth, our analyses included about twice as many lexical and semantic control factors as the prior studies, including multiple sources of word frequency estimates, and including the emotion × frequency interactions that are so important in word recognition. This greater stimulus control results in stronger internal validity for our study than for prior studies. Thus, overall, our results are more likely to be both internally and externally valid than prior results.

Theoretical Implications

The results also necessitate a new explanation of the affective effects in word processing. Previously, the automatic vigilance model was used to describe the origin of a valence effect that was thought to be categorical (Estes & Adelman, 2008a, 2008b), an inverted-U (Kousta et al., 2009), or interactive with arousal (Larsen et al., 2008). The present analyses revealed instead (1) that increasing valence speeds up lexical decisions, (2) that the effect is present across the entire range going from negative, over neutral, to positive words, (3) that the effect interacts with word frequency, and (4) that it does not interact with arousal (which itself has a small positive effect). The finding that the effect of valence is present across the entire continuum is a problem, for instance, for a view which attaches special status to negative (threatening) words, as this would predict a considerable difference between negative and neutral words but not between neutral and positive words. In fact, these results are problematic for all three of the prior models of automatic vigilance, as the effect of valence on RTs was neither categorical, an inverted-U, nor interactive with arousal. The present results instead suggest a gradient model of automatic vigilance, whereby a stimulus elicits a heightened effect in proportion to its negativity, and fine discriminations between negative, neutral and positive stimuli occur fast enough to influence the lexical decision or naming process.

Our results also reveal, for the first time, that arousal has a detrimental effect on word recognition times. More exciting words elicited slower responses. Among infrequent words this effect was about 40 ms in both lexical decision and naming, and again this effect was halved to about 20 ms among frequent words. The challenges are to explain why the effect (1) is detrimental, (2) is observed across the entire range, (3) interacts with word frequency, but (4) does not interact with valence. At the same time, it should be kept in mind that the contribution of arousal to lexical decision times is very small (.1% for the full dataset), so that it may not be warranted (yet) to come up with very strong theoretical proposals.

Factors influencing lexical decisions and naming can affect two processing stages: (1) the activation of word representations in the lexico-semantic system, and (2) the use of this information to execute a response (see Yap & Seow, 2013). Our correlational results do not allow us to pin down the sources of the effects, but plausible hypotheses do emerge from existing models of word recognition (e.g., Grainger & Jacobs, 1996; Norris, 2006) and affective priming (the finding that positive targets are processed faster after positive primes and negative targets faster after negative primes; e.g., Schmitz & Wentura, 2012; Spruyt, De Houwer, Hermans, & Eelen, 2007; Topolinski & Deutsch, 2013). These possible sources of the emotional effects on word processing are considered in detail below.

Lexico-semantic explanations of automatic vigilance—As Schmitz and Wentura (2012) report, there is a long-standing debate about the representation of valence in semantic memory. Bower (1991) suggested there were nodes for positive and negative valence in the semantic network with which valence-laden concepts were associated. In this way, the valence of concepts was not only known, but concepts (and hence words) could prime concepts of similar valence as well (i.e., affective priming). An alternative view was proposed by Masson (1995) and McRae, de Sa, and Seidenberg (1997). In their distributed

models, valence was coded in a series of units (roughly representing semantic features) and shared units between concepts made it easier to activate one concept on the basis of another. Topolinski and Deutsch (2013) showed that participants' affect changes briefly (for around 1 s) when stimuli with a strong positive or negative valence are presented, and critically for our purposes here, the degree of semantic priming is larger after positive affect inductions than after negative affect inductions. Thus, positive words may briefly lift the affect of the participants, increasing the affective or semantic priming of subsequent positive words. Negative words, in contrast, would temporarily induce negative affect and therefore prime responses to negative words, but crucially this negative affective priming would be smaller than positive affective priming.

Another possibility is that there are more positive word types than negative. A small but significant positivity bias is indeed observed in the rating study of Warriner et al. (2013), as 55.6% of about 14 thousand words were rated above the midpoint of the valence scale (5): positivity biases of a similar magnitude were also observed in multiple other corpora, see Kloumann et al. (2012) and references therein. Given that there are more positive words than negative words, more affective priming could occur for positive words than for negative words. Thus, positive words may elicit greater priming than neutral and negative words because (a) positive words are slightly more common (Warriner et al., 2013), and/or (b) positive words induce larger priming effects (Topolinski & Deutsch, 2013). That is, automatic vigilance could be due to affective priming, as positive words could produce more frequent or larger priming effects than negative words.

A lexico-semantic origin of the valence effect would also offer a parsimonious explanation of why the effect interacts with word frequency (see Kahan & Hely, 2008; Scott et al., 2009, 2012; Sheikh & Titone, in press for similar results in other tasks). Among less frequent words, the size of the valence effect was estimated by the regression model to be about 50 ms in lexical decisions and about 35 ms in naming (Figure 3). Among more frequent words, however, the effect of valence was reduced to about half that magnitude. This modulation by word frequency is common among lexico-semantic factors affecting word recognition. For instance, age of acquisition, letter-sound consistency, and imageability effects are also larger among low frequency words than among high frequency words (Cortese & Schock, 2013; Gerhand & Barry, 1999; Strain, Patterson, & Seidenberg, 1995). Typically, when two factors exert an interactive effect on word recognition, those factors are assumed to arise at the same stage of processing: If the two factors operated at different processing stages, it is unclear how they could interact. So given that frequency effects arise at the lexico-semantic stage of processing, and that frequency interacts with emotional factors, those emotional effects presumably also arise at the lexico-semantic stage.

Decision-response explanations of automatic vigilance—A second locus of the valence effect could be response execution (Yap & Seow, 2013). For instance, automatic vigilance could arise from task-specific processes. Much research in this respect has been done about the decision stage of the lexical decision task (see Kinoshita & Lupker, 2003, for context effects in naming). Two findings are particularly important: (1) lexical decisions are not always made on a full processing of the stimulus materials, and (2) any difference between word and nonword trials speeds up the decision process. Grainger and Jacobs

(1996) convincingly showed that "yes"-responses to words are partly based on the overall activation in the lexico-semantic system induced by the stimulus. That is, a yes-decision can be based on the fact that the stimulus activates many resembling word representations rather than on the identification of the stimulus itself. This explains why nonwords with many word neighbors elicit more erroneous responses than nonwords with few word neighbors, and why reaction times to words are faster when the nonwords do not resemble words than when they do (because then the overall activity elicited by the stimulus makes it possible to come to a correct decision). Within this view, positive words could result in faster responses because they have a lower response threshold, perhaps because positive stimuli are less life-threatening than negative stimuli and/or because humans in general seem to show a positivity bias in information processing (Walker, Skowronski, & Thompson, 2003). This would also explain why the valence effect is smaller (or even reversed) in participants with depression (Sharot, 2011) and when participants are brought into a situation that questions unrealistic optimism (Shepperd, Ouelette, & Fernandez, 1996).

Finally, it is simply possible that nonwords in general are perceived as slightly negative because they are unfamiliar: Warriner et al. (2013) show that lower-frequency words tend to be rated with lower valence. If this is the case, the valence of the stimulus will provide information about its "wordness" and will speed up the acceptance of positive words (e.g., Keuleers & Brysbaert, 2011). Thus, the automatic vigilance hypothesis – that negative stimuli engage attention longer than other stimuli – can be translated into "require more word-specific activation" or a "higher level of activation" to exceed the response threshold in a lexical decision task.

An explanation in terms of decision factors makes sense of seemingly contradictory results. Because negative stimuli in general require faster responses, they tend to be detected more rapidly. For instance, Nasrallah and colleagues (2009) subliminally presented negative, neutral, and positive words in an emotion detection task, and they found that negative words were identified more accurately than positive words. This finding suggests that negative stimuli are identified *faster*, or earlier, than other stimuli. However, the automatic vigilance hypothesis was developed to account for the observation that these same words in other tasks elicit *slower* responding (see also Pratto & John, 1991; Williams et al., 1996). A simple solution is that negative stimuli hold attention longer than other stimuli (Fox et al., 2001), and this sustained attention to negativity delays responding on other tasks such as color naming. After all, if the adaptive significance of automatic vigilance is to facilitate avoidance of dangerous stimuli, then negativity should speed rather than slow responding. Estes and Verges (2008) tested this hypothesis directly by having participants make either lexical decisions or valence judgments to the same set of negative words and positive words. Whereas the negative words slowed lexical decisions (as in the present study), they elicited faster valence judgments than positive words. Thus, automatic vigilance does not work by generally slowing responses to negative stimuli. Rather, by this account, negativity slows lexical decisions and color naming because valence is irrelevant to those judgments and therefore must be ignored or disengaged (cf. Fox et al., 2001; Kuperman, 2013).

An explanation in terms of decision factors also readily accounts for the finding that valence has a smaller effect on naming than on lexical decision, because the naming task is less

susceptible to decision processes (but see Kinoshita & Lupker, 2003, for evidence that it is not completely insusceptible to decisional factors). Whereas valence and arousal collectively explained 2% of the variance in lexical decision latencies, they explained only 0.3% of the variance in naming latencies.

Lexical processing—Finally, this research also contributes to our understanding of which variables affect performance in word processing tasks. Adelman et al. (2013) demonstrated that even after removing the random noise in word recognition times, the currently best-performing models and sets of word features leave unexplained a relatively large percentage of the variance in word recognition times. Similarly, although Rey and Courrieu (2010) noticed that there is 85% systematic variance in megastudy lexical decision data, current models do not go beyond 65% (e.g., Kuperman et al., 2012). Therefore Adelman et al. (2013) issued a general call to the field to search for additional factors that affect word recognition, and the present research does just that. Given the broad influence of emotion on cognitive tasks, it is rather surprising that current psycholinguistic models of word recognition entirely neglect the effects of emotion. Although valence and arousal exerted very modest effects on naming times (see also Adelman et al., 2013), we found that valence and arousal collectively explained a reasonably substantial amount (about 2%) of the unique variance in lexical decision times, with most of that effect arising from valence rather than arousal. Although this is a modest effect, it is a further step towards our understanding of which variables do and do not matter in language processing. For instance, it appears that valence may be a more important variable than many of the semantic richness variables recently proposed as relevant for characterizing word recognition.

Acknowledgments

References

Adelman, JS., editor. Visual word recognition, Volume 1: Models and methods, orthography and phonology. Psychology Press; Hove, England: 2012.

Adelman JS, Brown GDA, Quesada JF. Contextual diversity, not word frequency, determines word naming and lexical decision times. Psychological Science. 2006; 17(9):814–823. [PubMed: 16984300]

Adelman JS, Estes Z. Emotion and memory: A recognition advantage for positive and negative words independent of arousal. Cognition. 2013; 129:530–535. [PubMed: 24041838]

Adelman JS, Marquis SJ, Sabatos-DeVito MG, Estes Z. The unexplained nature of reading. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2013; 39:1037–1053.

Algom D, Chajut E, Lev S. A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. Journal of Experimental Psychology: General. 2004; 133(3):323–338. [PubMed: 15355142]

Baayen RH, Milin P. Analyzing reaction times. International Journal of Psychological Research. 2010; 3:12–28.

Balling L, Baayen R. Probability and surprisal in auditory comprehension of morphologically complex words. Cognition. 2012; 125:80–106. [PubMed: 22841290]

Balota DA, Cortese MJ, Sergent-Marshall S, Spieler DH, Yap MJ. Visual word recognition of single-syllable words. Journal of Experimental Psychology: General. 2004; 133(2):283–316. [PubMed: 15149254]

Balota DA, Lorch RF. Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1986; 12(3):336–345.

Balota, DA.; Yap, MJ.; Hutchison, KA.; Cortese, MJ. Megastudies: What do millions (or so) of trials tell us about lexical processing?. In: Adelman, JS., editor. Visual word recognition volume 1: Models and methods, orthography and phonology. Psychology Press; Hove, England: 2012.

Balota DA, Yap MJ, Hutchinson KA, Cortese MJ, Kessler B, Loftis B, Neely JH, Nelson DL, Simpson GB, Treiman R. The English Lexicon Project. Behavior Research Methods. 2007; 39(3):445–459. [PubMed: 17958156]

Barr D, Levy R, Scheepers C, Tily H. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013; 68(3):255–278.

Bower, GH. Mood congruity of social judgments.. In: Forgas, JP., editor. Emotion and social judgments. Pergamon; Elmsford, NY: 1991. p. 31-53.

Box G, Cox D. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological). 1964; 26(2):211–252.

Bradley, MM.; Lang, PJ. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report C-1. The Center for Research in Psychophysiology, University of Florida; 1999.

Brysbaert M, Buchmeier M, Conrad M, Jacobs AM, Bölte J, Böhl A. The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. Experimental Psychology. 2011; 58:412–424. [PubMed: 21768069]

Brysbaert M, Cortese MJ. Do the effects of subjective frequency and age of acquisition survive better word frequency norms? Quarterly Journal of Experimental Psychology. 2011; 64(3):545–559.

Brysbaert M, New B. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods. 2009; 41(4):977–990. [PubMed: 19897807]

Brysbaert M, New B, Keuleers E. Adding Part-of-Speech information to the SUBTLEX-US word frequencies. Behavior Research Methods. 2012; 44:991–997. [PubMed: 22396136]

Butler B, Hains S. Individual differences in word recognition latency. Memory & Cognition. 1979; 7(2):68–76.

Burgess C, Livesay K. The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kucera and Francis. Behavior Research Methods. 1998; 30(2):272–277.

Cortese M, Fugett A. Imageability ratings for 3,000 monosyllabic words. Behavior Research Methods. 2004; 36(3):384–387.

Cortese MJ, Khanna MM. Age of acquisition predicts naming and lexical decision performance above and beyond 22 other predictor variables: An analysis of 2342 words. Quarterly Journal of Experimental Psychology. 2007; 60(8):1072–1082.

Cortese MJ, Schock J. Imageability and age of acquisition effects in disyllabic word recognition. Quarterly Journal of Experimental Psychology. 2013 Advance online publication. doi: 10.1080/17470218.2012.722660.

Erdelyi MH. A new look at the New Look: Perceptual defense and vigilance. Psychological Review. 1974; 81(1):1–25. [PubMed: 4812878]

Estes Z, Adelman JS. Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). Emotion. 2008a; 8:441–444. [PubMed: 18729575]

Estes Z, Adelman JS. Automatic vigilance for negative words is categorical and general. Emotion. 2008b; 8:453–457.

Estes Z, Jones LL, Golonka S. Emotion affects similarity via social projection. Social Cognition. 2012; 30:582–607.

Estes Z, Verges M. Freeze or flee? Negative stimuli elicit selective responding. Cognition. 2008; 108:557–565. [PubMed: 18433742]

Faust ME, Balota DA, Spieler DH, Ferraro FR. Individual differences in information processing rate and amount: Implications for group differences in response latency. Psychological Bulletin. 1999; 125(6):777–799. [PubMed: 10589302]

Forgas JP. Mood and judgment: The affect infusion model (AIM). Psychological Bulletin. 1995; 117(1):39–66. [PubMed: 7870863]

Fox E, Russo B, Bowles R, Dutton K. Do threatening stimuli draw or hold visual attention in subclinical anxiety? Journal of Experimental Psychology: General. 2001; 130(4):681–700. [PubMed: 11757875]

Gerhand S, Barry C. Age of acquisition and frequency effects in speeded word naming. Cognition. 1999a; 73(2):B27–B36. [PubMed: 10580164]

Gerhand S, Barry C. Age of acquisition, word frequency, and the role of phonology in the lexical decision task. Memory & Cognition. 1999b; 27(4):592–602. [PubMed: 10479818]

Grainger J, Jacobs AM. Orthographic processing in visual word recognition: a multiple read-out model. Psychological review. 1996; 103(3):518–565. [PubMed: 8759046]

Hastie, T.; Tibshirani, R. Generalized additive models. Chapman & Hall; London: 1990.

Hoffman P, Ralph MAL, Rogers TT. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. Behavior research methods. 2012; 45:718–730. [PubMed: 23239067]

Jones MN, Johns BT, Recchia G. The role of semantic diversity in lexical organization. Canadian Journal of Experimental Psychology. 2012; 66:115–124. [PubMed: 22686160]

Juhasz BJ, Yap MJ. Sensory experience ratings for over 5,000 mono- and disyllabic words. Behavioral Research Methods. 2013; 45:160–168.

Juhasz BJ, Yap MJ, Dicke J, Taylor S, Gullick M. Tangible words are recognized faster: the grounding of meaning in sensory and perceptual systems. Quarterly Journal of Experimental Psychology. 2011; 64:1683–1691.

Kahan TA, Hely CD. The role of valence and frequency in the emotional Stroop task. Psychonomic Bulletin & Review. 2008; 15(5):956–960. [PubMed: 18926988]

Kensinger EA, Corkin S. Two routes to emotional memory: Distinct neural processes for valence and arousal. Proceedings of the National Academy of Sciences. 2004; 101(9):3310–3315.

Keuleers E, Brysbaert M. Detecting inherent bias in lexical decision experiments with the LD1NN algorithm. The Mental Lexicon. 2011; 6(1):34–52.

Keuleers E, Lacey P, Rastle K, Brysbaert M. The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. Behavior Research Methods. 2012; 44(1):287–304. [PubMed: 21720920]

Kinoshita S, Lupker SJ. Priming and attentional control of lexical and sublexical pathways in naming: A reevaluation. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2003; 29(3):405–415.

Kliegl R, Masson M, Richter E. A linear mixed model analysis of masked repetition priming. Visual Cognition. 2010; 18(5):655–681.

Kloumann IM, Danforth CM, Harris KD, Bliss CA, Dodds PS. Positivity of the English language. PloS one. 2012; 7(1):e29484. [PubMed: 22247779]

Koesling K, Kunter G, Baayen RH, Plag I. Prominence in triconstituent compounds: Pitch contours and linguistic theory. Language and Speech. Advance online publication. 2013 doi: 10.1177/0023830913478914.

Kousta S-T, Vinson DP, Vigliocco G. Emotion words, regardless of polarity, have a processing advantage over neutral words. Cognition. 2009; 112(3):473–481. [PubMed: 19591976]

Kryuchkova T, Tucker BV, Wurm L, Baayen RH. Danger and usefulness in auditory lexical processing: evidence from electroencephalography. Brain and Language. 2012; 122:81–91. [PubMed: 22726720]

Kuperman V. Accentuate the positive: Semantic access in English compounds. Frontiers in Language Sciences. 2013; 4(203):1–10.

Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods. 2012; 44:978–990. [PubMed: 22581493]

Kuperman V, Van Dyke JA. Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. Journal of Experimental Psychology: Human Perception and Performance. 2013; 39(3):802–823. [PubMed: 23339352]

LaBar KS, Cabeza R. Cognitive neuroscience of emotional memory. Nature Reviews Neuroscience. 2006; 7(1):54–64.

Lang PJ, Bradley MM, Cuthbert BN. Emotion, attention, and the startle reflex. Psychological Review. 1990; 97:377–395. [PubMed: 2200076]

Larsen RJ, Mercer KA, Balota DA. Lexical characteristics of words used in emotional Stroop experiments. Emotion. 2006; 6(1):62–72. [PubMed: 16637750]

Larsen RJ, Mercer KA, Balota DA, Strube MJ. Not all negative words slow down lexical decision and naming speed: Importance of word arousal. Emotion. 2008; 8(4):445–452.

Mason C, Perreault W Jr. Collinearity, power, and interpretation of multiple regression analysis. Journal of Marketing Research. 1991; 28(3):268–280.

Masson MEJ. A distributed memory model of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1995; 21:3–23.

McRae K, de Sa VR, Seidenberg MS. On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology: General. 1997; 126:99–130. [PubMed: 9163932]

Miller GA. WordNet: a lexical database for English. Communications of the ACM. 1995; 38(11):39–41.

Nasrallah M, Carmel D, Lavie N. "Murder she wrote": Enhanced sensitivity to negative word valence. Emotion. 2009; 9(5):609–618. [PubMed: 19803583]

Norris D. The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. Psychological Review. 2006; 113(2):327–357. [PubMed: 16637764]

Ohman A, Mineka S. Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychological Review. 2001; 108(3):483–522. [PubMed: 11488376]

Osgood, CE.; Suci, G.; Tannenbaum, P. The Measurement of Meaning. University of Illinois Press; Urbana, IL: 1957.

Pratto F, John OP. Automatic vigilance: The attention-grabbing power of negative social information. Journal of Personality and Social Psychology. 1991; 61(3):380–391. [PubMed: 1941510]

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2012 ISBN 3-900051-07-0.

Rey A, Courrieu P. Accounting for item variance in large-scale databases. Frontiers in Psychology. 2010; 1 doi: 10.3389/fpsyg.2010.00200.

Robinson M, D. Storbeck J, Meier BP, Kirkeby BS. Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. Personality and Social Psychology Bulletin. 2004; 30(11):1472–1484. [PubMed: 15448310]

Rowe G, Hirsh JB, Anderson AK. Positive affect increases the breadth of attentional selection. Proceedings of the National Academy of Sciences. 2007; 104:383–388.

Russell JA. Core affect and the psychological construction of emotion. Psychological Review. 2003; 110(1):145–172. [PubMed: 12529060]

Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. Journal of Personality and Social Psychology. 1999; 76(5):805–819. [PubMed: 10353204]

Schmitz M, Wentura D. Evaluative Priming of Naming and Semantic Categorization Responses Revisited: A Mutual Facilitation Explanation. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38:984–1000.

Schock J, Cortese M, Khanna M. Imageability estimates for 3,000 disyllabic words. Behavior Research Methods. 2012; 44(2):374–379. [PubMed: 22037976]

Scott GG, O'Donnell PJ, Leuthold H, Sereno SC. Early emotion word processing: Evidence from event-related potentials. Biological Psychology. 2009; 80(1):95–104. [PubMed: 18440691]

Scott G, O'Donnell P, Sereno S. Emotion words affect eye fixations during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38(3):783–792.

Sheikh NA, Titone DA. Sensorimotor and linguistic information attenuate emotional word processing benefits: An eye movement study. Emotion. in press.

Sharot T. The optimism bias. Current Biology. 2011; 21(23):R941–R945. [PubMed: 22153158]

Shepperd JA, Ouellette JA, Fernandez JK. Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. Journal of Personality and Social Psychology. 1996; 70(4):844–855.

Spruyt A, De Houwer J, Hermans D, Eelen P. Affective priming of nonaffective semantic categorization responses. Experimental Psychology. 2007; 54(1):44–53. [PubMed: 17341014]

Strain E, Patterson K, Seidenberg MS. Semantic effects in single-word naming. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1995; 21(5):1140–1154.

Tillotson S, Siakaluk P, Pexman P. Body-object interaction ratings for 1,618 monosyllabic nouns. Behavioral Research Methods. 2008; 40:1075–1078.

Topolinski S, Deutsch R. Phasic affective modulations of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2013; 39:414–436.

Tremblay, A.; Baayen, RH. Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall.. In: Wood, D., editor. Perspectives on Formulaic Language: Acquisition and communication. The Continuum International Publishing Group; London: 2010. p. 151-173.

van Kleef GA. How emotions regulate social life: The emotions as social information (EASI) model. Current Directions in Psychological Science. 2009; 18:184–188.

Vinson, D.; Ponari, M.; Vigliocco, G. How does emotional content affect lexical processing?. Proceedings of the Cognitive Science Society Conference; Berlin. 2013.

Walker WR, Skowronski JJ, Thompson CP. Life is pleasant--and memory helps to keep it that way!. Review of General Psychology. 2003; 7(2):203–210.

Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods. 2013 Advance Online Publication. doi: 10.3758/s13428-012-0314-x.

Wieling M, Nerbonne J, Baayen RH. Quantitative social dialectology: Explaining linguistic variation geographically and socially. PLoS ONE. 2011; 6(9):e23613. [PubMed: 21912639]

Wentura D, Rothermund K, Bak P. Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. Journal of Personality and Social Psychology. 2000; 78(6):1024–1037. [PubMed: 10870906]

Williams JM, Mathews A, MacLeod C. The emotional Stroop task and psychopathology. Psychological Bulletin. 1996; 120(1):3–24. [PubMed: 8711015]

Wood, S. Generalized Additive Models. Chapman & Hall/CRC; New York: 2006.

Wood S. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B). 2011; 73:3–36.

Yap MJ, Balota DA. Visual word recognition of multisyllabic words. Journal of Memory and Language. 2009; 60(4):502–529.

Yap MJ, Balota DA, Sibley DE, Ratcliff R. Individual differences in visual word recognition: Insights from the English Lexicon Project. Journal of Experimental Psychology: Human Pereception and Performance. 2012; 38(1):53–79.

Yap MJ, Seow CS. The influence of emotion on lexical processing: Insights from RT distributional analysis. Psychonomic Bulletin & Review. 2013 advance online publication.

Yarkoni T, Balota D, Yap M. Moving beyond Coltheart's N: A new measure of orthographic similarity. Psychonomic Bulletin & Review. 2008; 15(5):971–979. [PubMed: 18926991]

Zeno, S.; Ivens, SH.; Millard, RT.; Duvvuri, R. The educator's word frequency guide. Touchstone Applied Science Associates; 1995.
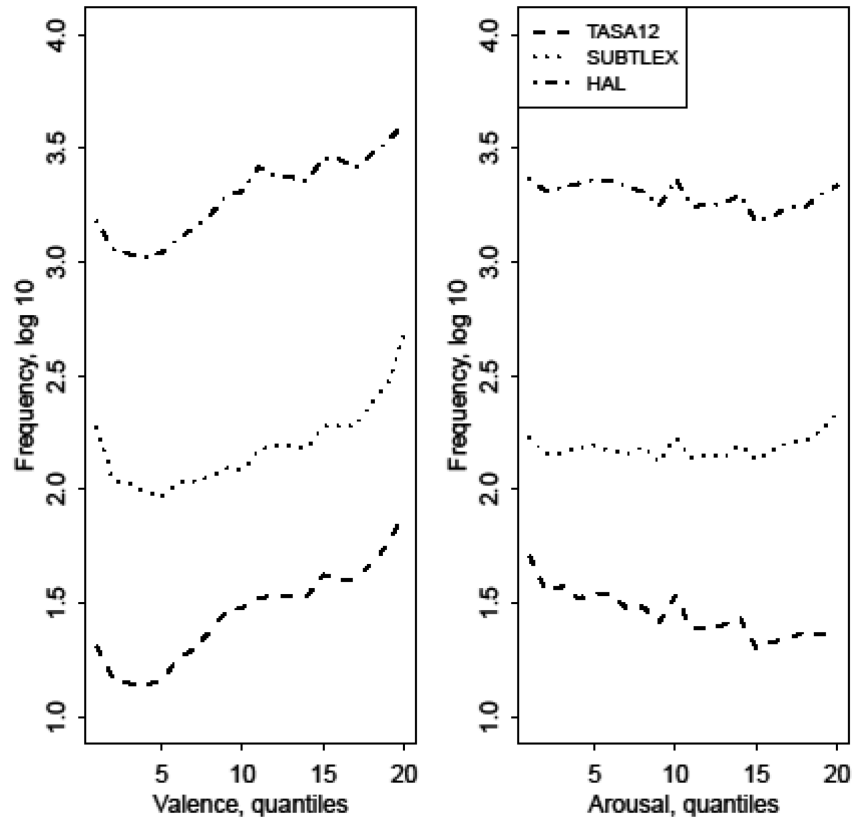
Figure 1.
Functional relation between valence and word frequency (left) and arousal and word frequency (right). Log (10)-transformed word frequencies are estimated for the SUBTLEX corpus based on subtitles to US films and media, the TASA12 corpus based on reading materials for American 12th graders, and the HAL corpus based on internet communications. Valence and arousal are binned into twenty 5% quantiles and the mean log frequency is shown for each quantile.
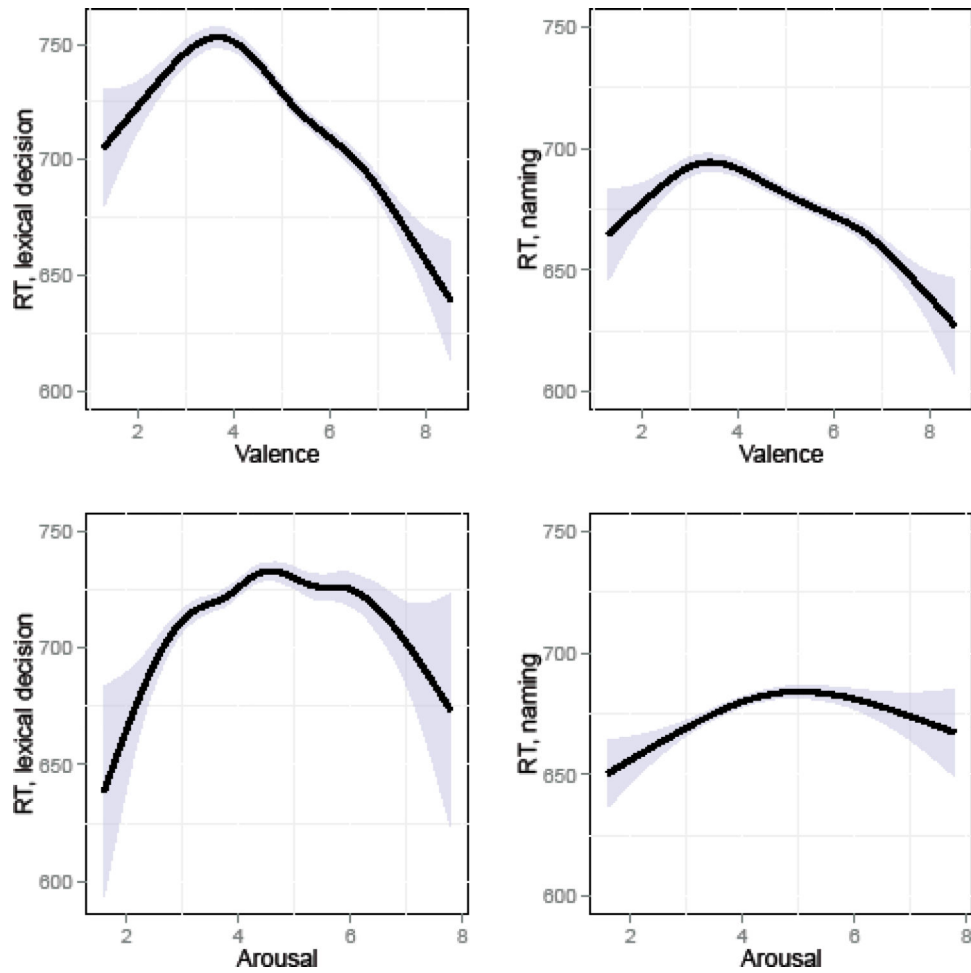
Figure 2.
Functional relationships of valence (top row) and arousal (bottom row) with lexical decision latencies (left column) and naming latencies (right column) across all frequency levels (i.e. emotion × frequency interactions are unaccounted for). The shape of valence and arousal effects was evaluated using cubic splines. Each curve is reported with the 95% confidence interval (the gray area).

Figure 3.
Functional relationships of valence (top row) and arousal (bottom row) with lexical decision latencies (left column) and naming latencies (right column), displayed by quintiles of word frequency (SUBTLEX). The highest-frequency words are the 5th quintile. The shape of valence and arousal effects was evaluated using cubic splines. Each curve is reported with the 95% confidence interval (the gray area).
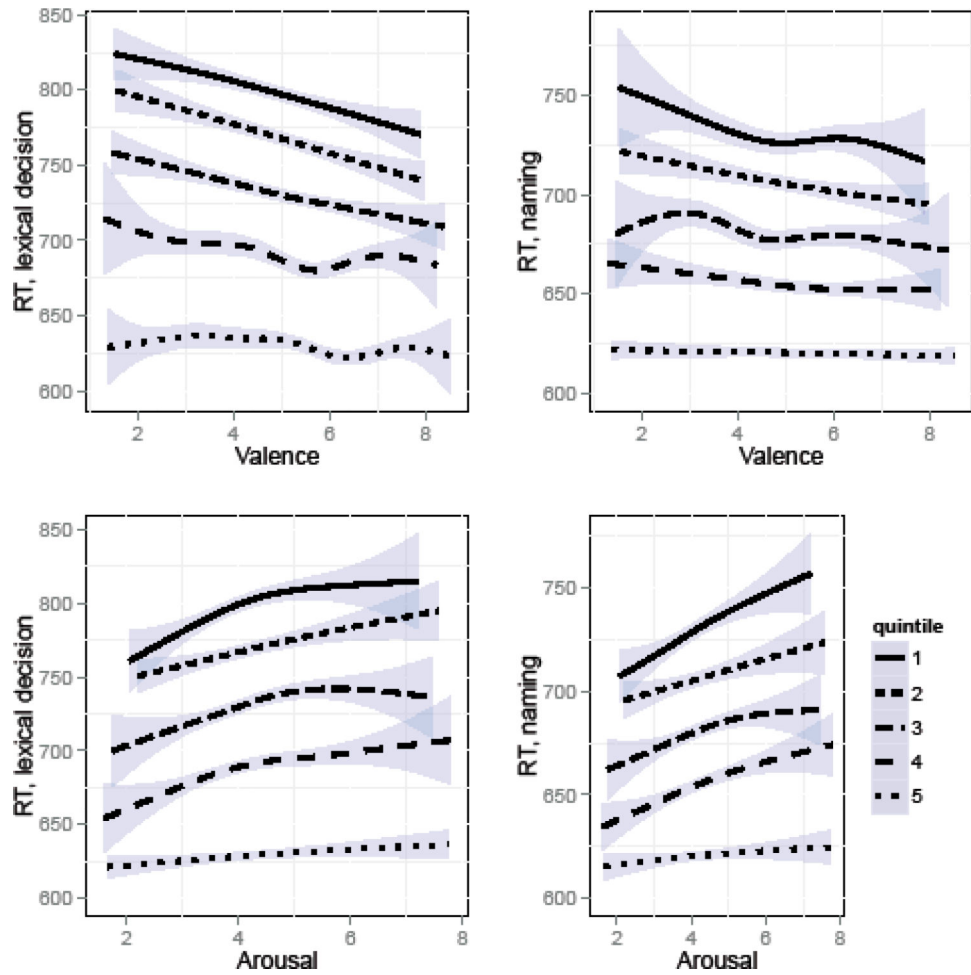
Table 1

Regression studies of the influence of emotion on word recognition latencies. EA (2008a) = Estes and Adelman (2008a); Kousta = Kousta et al. (2009); Lex Dec = lexical decision; *ns* = nonsignificant;

	EA (2008a)		Larsen et al. (2008)		Kousta
N	1011		1021		1446
	Lex Dec	**Naming**	**Lex Dec**	**Naming**	**Lex Dec**
Emotional Factors					
Arousal	***	**	*ns*	*ns*	*ns*
Valence	***	***	***	***	*
Arousal × Valence			***	*ns*	
Control Factors					
Letters	***	***	***	***	***
Syllables	*ns*	*			
Morphemes					*ns*
Frequency	**	*ns*	***	***	***
Familiarity					***
Contextual diversity	*ns*	*			
Orthographic N	*ns*	*ns*	*ns*	*ns*	***
Initial Phoneme		***			
Imageability					*ns*
Age of Acquisition					***
Bigram frequency					*ns*
Best R^2	53.24%	52.58%	58.70%	40.00%	64.55%

*
 $p < .05$;

**
 $p < .01$;

***
 $p < .001$.

Table 2

Generalized mixed additive model fitted to log-transformed lexical decision latencies. Linear effects (**Part A**) include linear predictors, whereas smooth terms (**Part B**) include nonlinear predictors and interactions.

A. Linear effects	Estimate	SE	t	p
Intercept	6.5702	0.0047	1390.7666	< 0.0001
PC2	−0.0016	0.0021	−0.7743	0.4388

B. Smooth terms		edf	Ref.df	F	p
PC1		6.5178	7.6971	244.5655	< 0.0001
PC3		5.0416	6.1925	38.5429	< 0.0001
Age of acquisition (residual)		4.8625	6.0297	128.1196	< 0.0001
Contextual diversity (residual)		5.2017	6.3745	52.2712	< 0.0001
Frequency × valence (tensor product)		12.0771	14.4453	39.9977	< 0.0001
Frequency × arousal (tensor product)		5.0539	20.0000	1.3966	< 0.0001

Note. *Part of speech* and *First phoneme* were both categorical predictors with multiple levels (4 and 32 respectively). For brevity we omit their inferential estimates from the model's output. *edf* = estimated degrees of freedom, *Ref.df* = reference degrees of freedom.

Table 3

Generalized mixed additive model fitted to log-transformed naming latencies. Linear effects (**Part A**) include linear predictors, whereas smooth terms (**Part B**) include nonlinear predictors and interactions.

A. Linear effects	Estimate	SE	*t*	*p*
Intercept	6.4860	0.0039	1657.4851	< 0.0001

B. Smooth terms		*edf*	*Ref.df*	*F*	*p*
PC1		5.6380	6.8167	251.5300	< 0.0001
PC2		6.5550	7.7254	2.9642	0.0030
PC3		6.6463	7.7718	43.1921	< 0.0001
Age of acquisition (residual)		5.9141	7.1296	143.4290	< 0.0001
Contextual diversity (residual)		3.1576	4.0278	18.9447	< 0.0001
Frequency × valence (tensor product)		7.4234	8.1773	25.6161	< 0.0001
Frequency × arousal (tensor product)		3.8260	20.0000	0.9284	0.0001

Note. *Part of speech* and *First phoneme* were both categorical predictors with multiple levels (4 and 32 respectively). For brevity we omit their inferential estimates from the model's output. *edf* = estimated degrees of freedom, *Ref.df* = reference degrees of freedom.

Table 4

Spearman's correlations of valence and arousal with semantic richness measures.

Measure	Valence	Arousal
Body-object interaction	.15[*]	−0.15[*]
Imageability	.19[*]	−.09[*]
Number of senses	.11[*]	.00
Semantic diversity	.07[*]	.00
Sensory experience	.07[*]	.19[*]

[*]
$p < .05$.

Table 5

Generalized Mixed Additive Model Fitted to Log-Transformed Naming Latencies: Smooth Terms Include Nonlinear Predictors and Interaction

Smooth terms	Estimated *df*	Reference *df*	*F*	*p*
PC1	5.6380	6.8167	251.5300	<0.0001
PC2	6.5550	7.7254	2.9642	0.0030
PC3	6.6463	7.7718	43.1921	<0.0001
Age of acquisition (residual)	5.9141	7.1296	143.4290	<0.0001
Contextual diversity (residual)	3.1576	4.0278	18.9447	<0.0001
Frequency Valence (tensor product)	7.4234	8.1773	25.6161	<0.0001
Frequency Arousal (tensor product)	3.8260	20.0000	0.9284	0.0001

Note. df = degrees of freedom; PC = principal component.