AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 12, 3566–3593
https://doi.org/10.1037/xge0001477

Emotion From the Sound of a Word: Statistical Relationships Between Surface Form and Valence of English Words Influence Lexical Access and Memory

Greig I. de Zubicaray[1], Katie L. McMahon[2, 3], Joanne Arciuli[4], Elaine Kearney[1], and Frank H. Guenther[5, 6]
[1] School of Psychology and Counselling, Faculty of Health, Queensland University of Technology (QUT)
[2] School of Clinical Sciences, Centre for Biomedical Technologies, Queensland University of Technology (QUT)
[3] Herston Imaging Research Facility, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia
[4] College of Nursing and Health Sciences, Flinders University
[5] Department of Speech, Language, and Hearing Sciences, Boston University
[6] Department of Biomedical Engineering, Boston University

It is generally accepted that a word's emotional valence (i.e., whether a word is perceived as positive, negative, or neutral) influences how it is accessed and remembered. There is also evidence that the affective content of some words is represented in nonarbitrary sound–meaning associations (i.e., emotional sound symbolism). We investigated whether more extensive statistical relationships exist between the surface form properties of English words and ratings of their emotional valence, that is, form typicality. We found significant form typicality for both valence and extremity of valence (the absolute distance from the midpoint of the rating scale, regardless of polarity). Next, using behavioral megastudy data sets, we show that measures of emotional form typicality are significant predictors of lexical access during written and auditory lexical decision and reading aloud tasks in addition to recognition memory performance. These findings show nonarbitrary form-valence mappings in English are accessed automatically during language and verbal memory processing. We discuss how these findings might be incorporated into theoretical accounts that implement Bayesian statistical inference.

Public Significance Statement

For over a century, language researchers have generally assumed that the relationship between the sound of a word and its meaning is entirely arbitrary. Our study shows there are systematic associations between the emotional valence of English words and their sound features. We also show that these emotional sound–meaning associations influence language processing and memory.

Keywords: form typicality, emotion, valence, word recognition, memory

We communicate emotion through language, even with a single word, and it is now generally accepted that a word's affective content influences how it is accessed and remembered. In addition, a word's affective content is not restricted to referencing an emotional state such as pleasure, grief, or fear. For example, the word "cancer" tends to be associated with negative emotional valence while "gift" tends to be considered positive, despite neither referring to a specific affective state (e.g., Bradley & Lang, 1999; S. Mohammad, 2018; Warriner et al., 2013). The observation that valence can be attributed across the vocabulary has been interpreted by some as reflecting grounding of emotional word meanings in bodily or interoceptive experience (e.g., Connell et al., 2018; Vigliocco et al., 2009). Yet, precisely how emotional valence influences word processing and memory remains unclear.

Much of the empirical work on emotional word processing has employed the lexical decision task (LDT) that requires participants

This article was published Online First September 14, 2023.
Greig I. de Zubicaray https://orcid.org/0000-0003-4506-0579
This research was supported by an Australian Research Council Discovery Project (Grant DP 220101853). The authors are grateful to Marc Brysbaert for his helpful comments on an earlier version of this work. All data and analysis scripts are publicly available at https://osf.io/q3uv7/.

Greig I. de Zubicaray served as lead for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, writing–original draft, and writing–review and editing. Katie L. McMahon served in a supporting role for conceptualization, funding acquisition, writing–original draft, and writing–review and editing. Joanne Arciuli served in a supporting role for funding acquisition, writing–original draft, and writing–review and editing. Elaine Kearney contributed equally to data curation and served in a supporting role for conceptualization, formal analysis, methodology, writing–original draft, and writing–review and editing. Frank H. Guenther served in a supporting role for conceptualization, funding acquisition, writing–original draft, and writing–review and editing. Katie L. McMahon and Elaine Kearney contributed equally to project administration.

Correspondence concerning this article should be addressed to Greig I. de Zubicaray, School of Psychology and Counselling, Faculty of Health, Queensland University of Technology (QUT), 149 Victoria Park Road, Kelvin Grove, QLD 4059, Australia. Email: greig.dezubicaray@qut.edu.au

to respond whether a written or spoken item is a word or nonword.[1] For written words, these studies have variously reported slower responses for negative versus positive and neutral words (e.g., Estes & Adelman, 2008a, 2008b; Kuperman et al., 2014; Larsen et al., 2008), faster responses for positive compared to neutral and negative words (e.g., Goh et al., 2016; Scott et al., 2014; Tucker et al., 2019), and faster responses for both positive and negative versus neutral words (i.e., an inverted U-shaped or quadratic relationship; Kousta et al., 2009; Siakaluk et al., 2016; Vinson et al., 2014; Yap & Seow, 2014). Using a bootstrapping regression approach with megastudy data sets, Kuperman (2015) subsequently provided evidence for both linear and quadratic relationships between emotional valence and written LDT latencies. Further complicating matters, there is evidence that the effect of valence might differ according to the behavioral megastudy data set employed for written word recognition (e.g., Gao, Shinkareva, & Peelen, 2022; Kuperman, 2015). The evidence from spoken word recognition is just as mixed, with some studies reporting that positive valence was associated with quicker responses in auditory LDT (a monotonic function; Tucker et al., 2019), or that both negative and positive words facilitated performance (an inverted-U function commensurate with extremity of valence; Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022) or both patterns (Goh et al., 2016).

All three directions of effect have been interpreted as confirming theoretical accounts prioritizing the different aspects of emotional valence in information processing. For example, the finding that negative words are responded to more slowly (e.g., Estes & Adelman, 2008a, 2008b) has been considered consistent with the automatic vigilance hypothesis (e.g., Pratto & John, 1991) that proposes negative stimuli are attended to preferentially, extending their processing time. An overall advantage for emotional words (e.g., Vinson et al., 2014) has been interpreted as supporting a model in which affective stimuli, regardless of polarity, have greater motivational relevance (e.g., Lang et al., 1990). Finally, the finding that positive words have a processing advantage (e.g., Scott et al., 2014) has been considered consistent with the proposal that positive information is processed more quickly because it is more densely clustered in semantic space (e.g., Unkelbach et al., 2008).

Some theoretical accounts explicitly propose a memory advantage for negative words. For example, the negative emotional valence enhances recapitulation (NEVER) model (Bowen et al., 2018) proposes that negatively valenced stimuli involve increased encoding of sensory detail and so are better integrated into, and reactivated from, memory. Others propose an overall memory advantage for emotional compared to neutral stimuli (e.g., Phelps & Sharot, 2008). The empirical literature on recognition memory and emotional valence is mixed, with almost as many studies reporting null results as those reporting an advantage over neutral words (for a review, see MacMillan et al., 2022). Adelman and Estes (2013) reported a recognition memory advantage for a measure of extremity of valence (i.e., the absolute distance from the midpoint of the valence rating scale, regardless of polarity, corresponding to the quadratic term for valence), but no effect for either positive or negative valenced words compared to neutral words once a range of lexical variables and arousal were controlled. This result was recently replicated by Lau et al. (2018) and Cortese and Khanna (2022).

Nonarbitrary Relationships Between Emotional Valence and Sound in English Words

It is well known that prosody plays an important role in communicating emotion (e.g., Frick, 1985; Huttar, 1968; Kamiloğlu et al., 2020) and that suprasegmental features can convey affective content independent of word meaning (e.g., S. K. Kim & Sumner, 2017). However, the emotional valence of words is also proposed to be represented in nonarbitrary sound–meaning associations across many languages including English (e.g., Adelman et al., 2018; Louwerse & Qu, 2017; Sidhu & Pexman, 2018). These include symbolic uses of articulation: height/space and frequency/intonation (i.e., fundamental frequency, f_o). Early studies of emotional sound symbolism reported correlations between certain vowels and valence in small sets of English words (e.g., Heise, 1966) or demonstrated participants were more likely to rate nonwords as pleasant or unpleasant if they contained certain vowel sounds (e.g., Taylor & Taylor, 1962).

Yu et al. (2021) recently reported that the vowels /iː/ and /ʌ/ in the middle phoneme position are significantly associated with positive and negative emotional states, respectively (the "gleam/glum effect"). They proposed the gleam-glum effect is grounded in congruent facial musculature patterns for expressing positive (e.g., smiling) and negative emotions (e.g., grimacing; Yu et al., 2021). However, the mean valence ratings of Yu et al.'s (2021) /iː/ and /ʌ/ words were 5.12 and 4.83, respectively (the mean difference of 0.29 was statistically significant) indicating both comprised words from the neutral midpoint of the 9-point rating scale from the norms they employed (Warriner et al., 2013).

Consonant articulation has also been linked to valence in English, with words whose consonants conform to inward wandering patterns (i.e., from front to back of mouth) proposed to be more positive (e.g., *strong*) than those with outward wanderings (e.g., *claim*; Bakhtiari et al., 2016; Topolinski et al., 2014). This "in–out effect" is proposed to be grounded in oral muscle contractions associated with ingestion and expectoration of food reflecting the motivational states of approach and avoidance, respectively (e.g., Topolinski et al., 2014; but see Maschmann et al., 2020). However, Engelen (2022) was recently unable to find an association between consonantal wandering direction and valence ratings (Warriner et al., 2013) to support the validity of the in–out effect in English.

Louwerse and Qu (2017) reported that nasal consonants (e.g., /n/, /m/) in the beginning of English words predicted negative valence according to the Warriner et al. (2013) norms. However, Adelman et al. (2018) argued that Louwerse and Qu's (2017) findings were largely attributable to negating prefixes such as "in-," "un,-" and "im-" as in *inedible*, *unhappy*, or *immature* rather than nasals in the first phoneme position as in *nasty* or *murder*. They instead reported that the initial phoneme was a significant predictor of valence and proposed that "front-loading" was a distinctive property of emotional sound symbolism in English occurring at the level of many individual phonemes rather than phonetic features. Furthermore, they proposed that emotional sound symbolism has a

[1] Early studies that reported effects of valence on word recognition and memory often failed to control for important lexical variables (see Estes & Adelman, 2008a, Larsen et al., 2008). Here we restrict our literature review to more recent investigations of independent effects of valence that controlled for these variables.

negative priority, with phonemes that are pronounced more rapidly tending to occur at the beginning of negative words. Adelman et al. (2018) hypothesized the reason why negative words might have initial phonemes that can be uttered more quickly is because "avoiding negative outcomes is more urgent than obtaining positive outcomes" (p. 127). However, several studies have reported that negative words are articulated more slowly than neutral words or that positive words are articulated more quickly (e.g., Algom et al., 2004; Estes & Adelman, 2008a, 2008b; Kuperman et al., 2014). To reconcile these findings with their automatic vigilance hypothesis, Adelman et al. argued that although negative phonemes are uttered more quickly, negative words may only be produced rapidly in contexts in which valence is relevant to the task, such as alarm signaling.

Aside from symbolic relationships, there are other nonarbitrary associations between word form and meaning that manifest more extensively within vocabularies as statistical regularities (Dingemanse et al., 2015). Corpus studies have demonstrated significant form systematicity or typicality in assignment of lexical stress, syllable duration, and vowel properties of nouns and verbs in English. In addition, these probabilistic phonological cues to syntactic category have been shown to predict performance on a range of tasks including grammatical category judgment, lexical decision, and articulation/pronunciation (e.g., Arciuli & Cupples, 2006, 2007; Arciuli & Monaghan, 2009; Cassidy & Kelly, 1991; G. I. de Zubicaray et al., 2021; de Zubicaray, Arciuli, Guenther, et al., 2023; Kelly, 1992; Monaghan et al., 2005, 2010; J. A. Sereno & Jongman, 1990; Sharpe & Marantz, 2017). Form typicality for concrete versus abstract concepts has also been demonstrated within English nouns (e.g., Reilly & Kean, 2007; Reilly et al., 2017; Westbury & Moroschan, 2009). More recently, systematic form–meaning mappings for ratings of sensory experience (the extent to which a word evokes a sensory and/or perceptual experience; Juhasz & Yap, 2013) and semantic size (a measure of magnitude of either concrete or abstract words; S. C. Sereno et al., 2009) have also been demonstrated, providing the first evidence that form typicality for these semantic variables also influences language processing and memory (e.g., G. de Zubicaray, Arciuli, Kearney, et al., 2023; G. I. de Zubicaray, Kearney, Arciuli, et al., 2023). The above results have been interpreted as indicating language users are able to exploit statistical regularities in form–meaning mappings automatically during language processing (e.g., Arciuli, 2017, 2018).

The Present Study

Precisely how the affective content of words influences their processing and memory remains unclear. The evidence for effects of valence in lexical decision and recognition memory is often weak and contradictory (see literature reviewed above). It is possible that nonarbitrary mappings between surface form features and emotional valence of English words influence how they are accessed. Yet, there is only limited evidence from sound symbolism of associations between some phonemes and valence (e.g., Adelman et al., 2018; Louwerse & Qu, 2017; Winter, 2023). In addition, form typicality is yet to be investigated more extensively within the English vocabulary in relation to lexical-semantic categories.

A possible reason for some of the mixed results in both literatures could be the indiscriminate use of subjective ratings of words' emotional valence. Pollock (2018) elegantly demonstrated the problem of using average ratings when participants' judgments disagree, noting a standard deviation (*SD*) above 1.5 "means that some people report a very strong negative response to that word, whereas some people report little or no emotional response at all. So if a researcher is interested in comparing responses to neutral words with responses to emotionally valenced words, they should definitely avoid words with high standard deviations for emotional valence, because they will add a significant amount of noise to the experimental design" (p. 1212). It is worth emphasizing that this issue affects all behavioral megastudies of valence effects in lexical decision and recognition memory using the Warriner et al. normative data set (e.g., Cortese & Khanna, 2022; Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman et al., 2014; Lau et al., 2018; Tucker et al., 2019; Vinson et al., 2014). It also affects all corpus investigations of emotional sound symbolism (e.g., Adelman et al., 2018; Yu et al., 2021) excepting possibly that of Louwerse and Qu (2017) who restricted their analyses to a subset of words with extreme values less likely to have rating disagreement. However, the use of extreme subsets in experimental designs also runs the risk of revealing a relationship that is not present in the full vocabulary (e.g., Keuleers & Balota, 2015; Kuperman, 2015; Liben-Nowell et al., 2019).

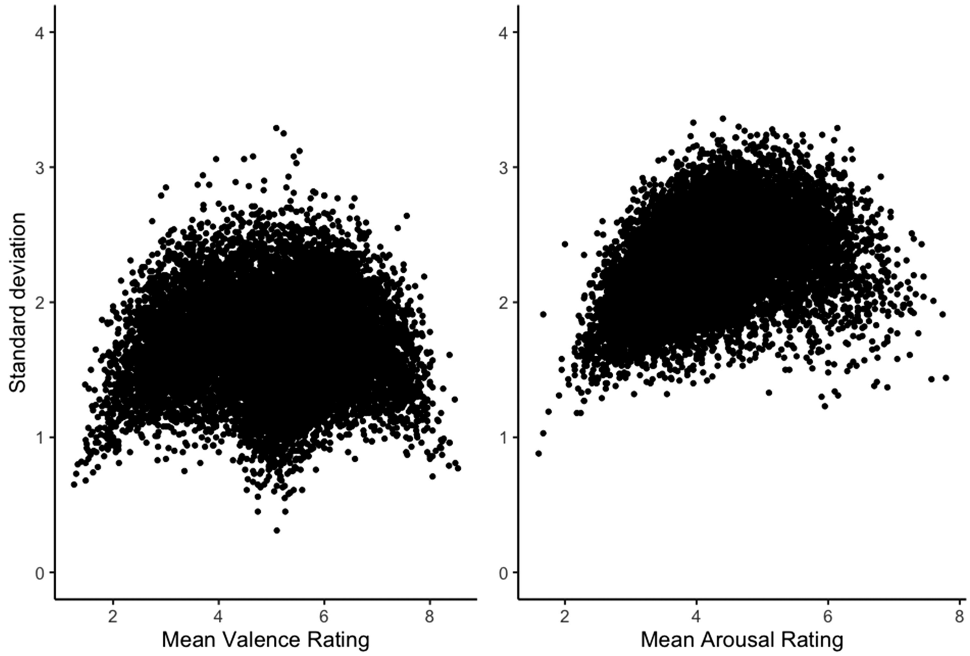
The aims of the present study were therefore twofold: (a) to determine the extent to which valence ratings of English words can be predicted using a large set of surface form features and (b) to determine whether a measure of form typicality for emotional valence can predict word processing and memory performance. To achieve the latter, we employed data sets for written lexical decision and articulation/reading aloud from the English Lexicon Project (ELP; Balota et al., 2007), written word recognition from the English Crowdsourcing Project (ECP; Mandera et al., 2020), auditory lexical decision from both the Auditory English Lexicon Project (AELP; Goh et al., 2020) and Massive Auditory Lexical Decision (MALD) project (Tucker et al., 2019), and recognition memory from Cortese and Khanna (2022). For the latter investigations, we conducted a series of hierarchical regression analyses, entering a range of lexical and semantic control variables first, followed by valence ratings and form typicality.

Study 1a: Investigating Form-Valence Mappings of English Words

The aim of Study 1a was to investigate systematic word form associations for valence and extremity of valence (Adelman & Estes, 2013) using 13,915 monosyllabic and multisyllabic words from Warriner et al.'s (2013) norms in which valence and arousal are assigned on a 9-point scale ranging from 1 (*unhappy/calm*) to 9 (*happy/excited*), respectively. Warriner et al. (2013) noted that the average *SD* of their valence ratings was 1.68. Figure 1 plots the mean valence and arousal values and the corresponding *SD* of every word rated in the Warriner et al. (2013) norms. It is evident from the plot that participants disagreed about the emotional valence of most words in the norms, as only 4,498 (32%) have *SD*s less than or equal to 1.5. As the plot also shows, the situation is far worse for arousal ratings, which Warriner et al. noted had an average *SD* of 2.30. Clearly, their participants disagreed about the arousal values of the overwhelming majority of words, with only 86 having *SD*s less than or equal to 1.5 (0.6%). Furthermore, there are only 30 words that have both arousal and valence rating *SD*s less than or equal to 1.5.

Figure 1

Valence and Arousal Ratings Agreement in the Warriner et al. (2013) Norms (N = 13,915)



The extent of the rating disagreement for arousal raises questions both about its validity as a measure of a word's affective content and the findings of studies that included it in analyses when investigating emotional sound symbolism (e.g., Adelman et al., 2018) and the impact of valence and extremity of valence on language and memory processing (e.g., Estes & Adelman, 2008a, 2008b; Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman et al., 2014). However, the extent of disagreement is perhaps not surprising, as Kuppens et al. (2013) showed that the relationships between multiple types of valence and arousal measures at the nomothetic level (i.e., averaged across participants) are rarely consistent with those at the idiographic level (i.e., for individuals), leading them to conclude "there is no universal architecture that adequately describes the relation between valence and arousal in subjective experience for everyone" (p. 934). Hence, for Study 1a, we followed Pollock's (2018) recommendation and selected only the subset of words with valence ratings showing reasonable agreement, excluding the arousal ratings completely.

Method

Materials

The initial data set comprised 4,498 English monosyllabic and multisyllabic words with *SD*s less than or equal to 1.5 from the Warriner et al. (2013) valence norms. We excluded items corresponding to proper names based on their Part of Speech classification (Brysbaert et al., 2012). For each word, we calculated 60 form variables, all of which were phonological/phonetic in nature excepting orthographic length, following Sharpe and Marantz's (2017) approach. We used phonemic transcriptions from the Carnegie Mellon University pronouncing dictionary along with stress category assignments from Tucker et al.'s (2019) database of 26,793 words. For each word, we coded whole word properties (length in letters, phonemes, and syllables), initial and final phonemes, number of and initial and final positions for typical phonetic features (i.e., place and manner of articulation for consonants; place and height for vowels; voicing), and syllable position for primary lexical stress: initial, final, medial (i.e., primary stress in the interior syllables), and multiple (i.e., more than one syllable with primary stress as some words have even stress across syllables). This resulted in a final set of 3,380 words common to all databases (727 monosyllabic, 1,357 disyllabic, 802 trisyllabic, 376 quadrisyllabic, 108 pentasyllabic, eight sexisyllabic, and two heptasyllabic). Extremity of valence was calculated by subtracting the midpoint of the scale (i.e., 5) from the Warriner et al. (2013) valence rating of each word and disregarding polarity.

Design and Analysis

All analyses were performed in R Version 4.2.1 (R Core Team, 2022). The form variables were first evaluated for linear dependencies using the *caret* package (*findLinearCombos*; Kuhn, 2022), resulting in the exclusion of the following variables: *number low*, *number glottal*, *number glide*, *number voiceless*, *first phoneme is glide*, *first phoneme is voiceless*, *final phoneme is glottal*, *final phoneme is liquid*, *final phoneme is glide*, and *final phoneme is voiceless*. Next, to determine the best subset of form variables for predicting valence and extremity of valence values, we used the *leaps* package (Lumley, 2022). To prevent overfitting, we selected the best-fit model for each in terms of its predictive accuracy via 10-fold cross-validation (repeated 200 times with different randomized folds) using the *caret* package (see de Rooij & Weeda, 2020; Yarkoni & Westfall, 2017), consistent with our intention to provide a set of form typicality values for future studies to explore. Models that minimize root-mean-square error for prediction are better in terms of both goodness of fit and complexity. We used 90% of the data as training and 10% as test. The best subsets of form variables for valence and extremity of valence, respectively, were then entered into separate hierarchical linear regressions with robust standard errors (*SE*s; Wilcox, 2016). Form typicality was calculated as the predicted value of the dependent variable for each word according to the robust model (e.g., Valence Form Typicality = Valence + Residual; e.g., Sharpe & Marantz, 2017). The values for all words were then *Z*-transformed. Hence, for valence ratings, positive typicality values indicate forms with positive valence and negative values forms with negative valence. For extremity of valence ratings, positive typicality values indicate forms that are highly emotionally valenced regardless of polarity, while negative values indicate more neutral forms.

To determine if our measures of form typicality for valence and extremity of valence differed according to the major lexical categories (adjectives, adverbs, nouns, and verbs; $N = 3{,}380$),[2] we conducted analyses of variance (ANOVAs). Bartlett's test showed the valence form typicality data violated the assumption of homogeneity of variance, $\chi^2(3) = 9.657$, $p = .022$, as did the extremity of valence data, $\chi^2(3) = 12.295$, $p = .006$. We, therefore, conducted Welch's ANOVAs, followed by Games–Howell post hoc tests using the package *rstatix* (Kassambara, 2021). Violin plots were generated using the package *ggstatsplot* (Patil, 2021).

Transparency and Openness

All data and analysis scripts for this and the subsequent regression studies are available for replication at: https://osf.io/q3uv7/.

Results and Discussion

The best-fit model for valence comprised 13 form variables, giving an adjusted R^2 of 0.027 ($p < .001$, Table 1). The best-fit model for extremity of valence comprised 16 form variables, giving an adjusted R^2 of 0.040 ($p < .001$, Table 2).

Figures 2 and 3 show form typicality as a function of part of speech. Typicality for valence differed significantly according to lexical category, Welch's $F(3, 10.04) = 4.915$, $p = .024$, est. $\omega^2 = .000$. However, post hoc Games–Howell tests revealed no significant differences (all *p*s > .05). By contrast, typicality for extremity of valence differed significantly, Welch's $F(3, 9.96) = 48.269$, $p < .001$, est. $\omega^2 = .048$. Post hoc Games–Howell tests revealed adjectives were significantly more typical forms than nouns ($M_{\text{diff}} = 0.484$, $p < .001$) and verbs ($M_{\text{diff}} = 0.585$, $p < .001$). All other comparisons were not significant (all *p*s > .05).

These results provide the first evidence of significant systematic form–meaning mappings for valence and extremity of valence in English words. Form variables explained 2.7% of variance in valence and slightly more variance (4.0%) in extremity of valence. In addition, although form typicality for valence did not differ significantly according to lexical category, adjectives were on average more typical forms for extremity of valence. Of note, we did not replicate Adelman et al.'s (2018) finding of "front loading," that is, the first phoneme significantly predicting valence, that they proposed to be a relatively distinctive property of emotional sound symbolism in English. However, we found multiple phonetic features of the first phoneme were able to predict valence significantly. This included replicating Louwerse and Qu's (2017) finding that nasals in the first phoneme predict negative valence in English. We also replicated their findings for labiodentals in words and velars in the first phoneme. However, Adelman et al. (2018) proposed that Louwerse and Qu's (2017) finding that initial nasals are negative is attributable to negating prefixes in English words such as "in-" and "un-" as in *inedible* or *unhappy*, which might also be the case here. Phonetic features were also strong predictors of extremity of valence. Surprisingly, nasals were also a significant predictor of more neutral words.

To determine whether our findings were influenced by morphophonological redundancies, we conducted Study 1b repeating the analyses with only monomorphemic words that lack affixes.

Table 1
Best-Fit Model for Predicting Valence With Form Variables According to 10-Fold Cross-Validation Repeated 200 Times (n = 3,380)

Model comparison	Estimate	*SE*	*T*
Intercept	4.958	0.111	44.863***
Number central	0.063	0.028	2.296*
Number labiodental	0.341	0.071	4.802***
Number stop	−0.076	0.029	−2.593**
Number fricative	−0.194	0.041	−4.748***
Final phoneme	0.006	0.003	1.826
First phoneme bilabial	0.457	0.108	4.222***
First phoneme alveolar	0.212	0.084	2.538*
First phoneme palatal	0.307	0.097	3.150**
First phoneme velar	0.653	0.118	5.552***
First phoneme stop	−0.295	0.088	−3.351***
First phoneme nasal	−0.400	0.124	−3.218**
First phoneme voiced	−0.170	0.062	−2.738**
Initial stress position	0.139	0.059	2.378*

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

Study 1b: Investigating Form-Valence Mappings of Monomorphemic English Words

The aim of Study 1b was to investigate form–meaning associations for valence in monomorphemic words. Otherwise, Study 1b was identical to Study 1a.

Method

Materials

The initial data set comprised 3,380 words with good rating agreement from the Warriner et al. (2013) norms employed in Study 1a. We then selected only monomorphemic words according to Tucker et al.'s (2019) database ($N = 2{,}186$; 700 monosyllabic, 903 disyllabic, 417 trisyllabic, 142 quadrisyllabic, 22 pentasyllabic, one sexisyllabic, and one heptasyllabic).

Design and Analysis

Identical to Study 1a. Bartlett's test showed the valence form typicality data violated the assumption of homogeneity of variance, $\chi^2(3) = 24.560$, $p < .001$. We, therefore, conducted Welch's

[2] As there was only one function word, this category was omitted from analyses.

Table 2

Best-Fit Model for Predicting Extremity of Valence With Form Variables According to 10-Fold Cross-Validation Repeated 200 Times (n = 3,280)

Model comparison	Estimate	*SE*	*t*
Intercept	0.937	0.051	18.232***
Number high	0.109	0.021	5.076***
Number mid	0.057	0.017	3.318***
Number stop	−0.051	0.020	−2.480*
Number fricative	0.129	0.022	5.728***
Number liquid	0.062	0.024	2.579**
First phoneme bilabial	−0.088	0.043	−2.047*
First phoneme palatal	−0.243	0.074	−3.270**
First phoneme glottal	0.203	0.098	2.082*
First phoneme stop	−0.109	0.046	−2.378*
First phoneme affricate	0.253	0.154	1.641
First phoneme fricative	−0.205	0.051	−4.053***
First phoneme nasal	−0.272	0.076	−3.565***
First phoneme liquid	−0.225	0.089	−2.524**
First phoneme voiced	0.165	0.042	3.936***
Final phoneme velar	0.161	0.056	2.903**
Final phoneme stop	0.085	0.040	2.135*

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

ANOVAs, followed by Games–Howell post hoc tests using the package *rstatix* (Kassambara, 2021).

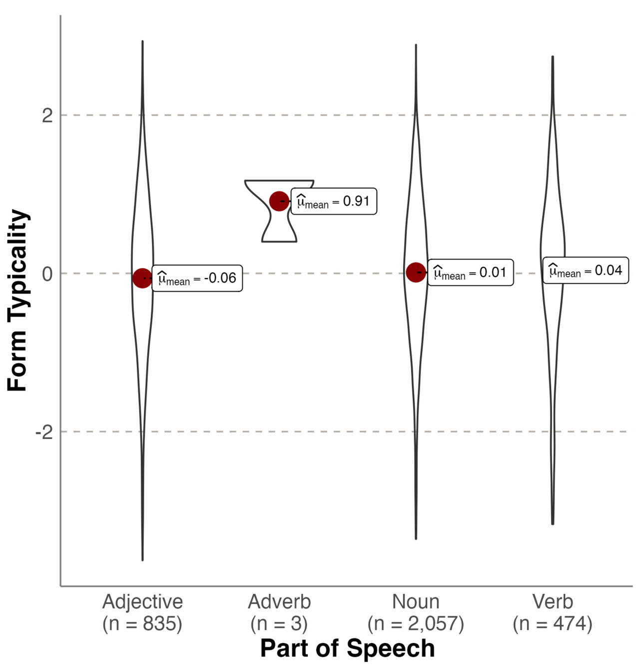
Results and Discussion

The best-fit model for valence comprised 12 form variables, giving an adjusted R^2 of 0.020 (Table 3). The best-fit model for extremity of valence comprised nine form variables, giving an adjusted R^2 of 0.013 (Table 4).

Figures 4 and 5 show form typicality as a function of part of speech. Typicality for valence again approached but did not reach significance for differing according to lexical category, Welch's $F(3, 10.10) = 3.345$, $p = .063$, est. $\omega^2 = .003$. Post hoc Games–Howell tests revealed adjectives were significantly more typical forms than verbs ($M_{\text{diff}} = 0.228$, $p = .022$). All other comparisons

Figure 2

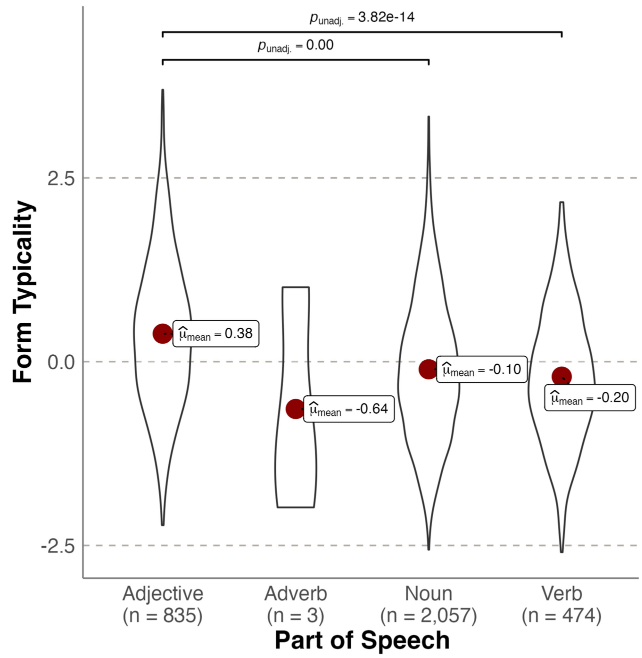
Violin Plot Showing Probability Densities of Valence Form Typicality Standardized Values as a Function of Lexical Category



Note. The shaded dot indicates the mean. See the online article for the color version of this figure.

Figure 3

Violin Plot Showing Probability Densities of Extremity of Valence Form Typicality Standardized Values as a Function of Lexical Category



Note. The shaded dot indicates the mean. p_{unadj} = unadjusted *p* value. See the online article for the color version of this figure.

Table 3

Best-Fit Model for Predicting Valence With Form Variables in Monomorphemic Words According to 10-Fold Cross-Validation Repeated 200 Times (n = 2,186)

Model comparison	Estimate	*SE*	*t*
Intercept	5.755	0.274	21.027***
Number stop	−0.103	0.037	−2.803**
Number fricative	−0.107	0.045	−2.374*
Final phoneme	0.007	0.004	2.047*
First phoneme bilabial	0.272	0.090	3.022**
First phoneme velar	0.443	0.106	4.176***
First phoneme stop	−0.133	0.092	−1.453
First phoneme nasal	−0.416	0.125	−3.334**
Final phoneme labiodental	0.418	0.166	2.517*
Number syllables	0.155	0.043	3.620***
Initial stress position	−0.861	0.246	−3.495***
Medial stress position	−1.062	0.262	−4.056***
Final stress position	−0.998	0.258	−3.868***

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

Table 4

Best-Fit Model for Predicting Extremity of Valence With Form Variables in Monomorphemic Words According to 10-Fold Cross-Validation Repeated 200 Times (n = 2,186)

Model comparison	Estimate	*SE*	*t*
Intercept	0.848	0.076	11.223***
Number high	0.052	0.029	1.801
Number bilabial	−0.047	0.028	−1.702
Number stop	−0.036	0.022	1.672
Final phoneme	0.004	0.002	1.555
First phoneme glottal	0.205	0.109	1.881
First phoneme nasal	−0.186	0.075	−2.479*
First phoneme voiced	0.087	0.042	2.099*
Final phoneme affricate	0.162	0.087	−1.865
Number syllables	0.062	0.023	2.743**

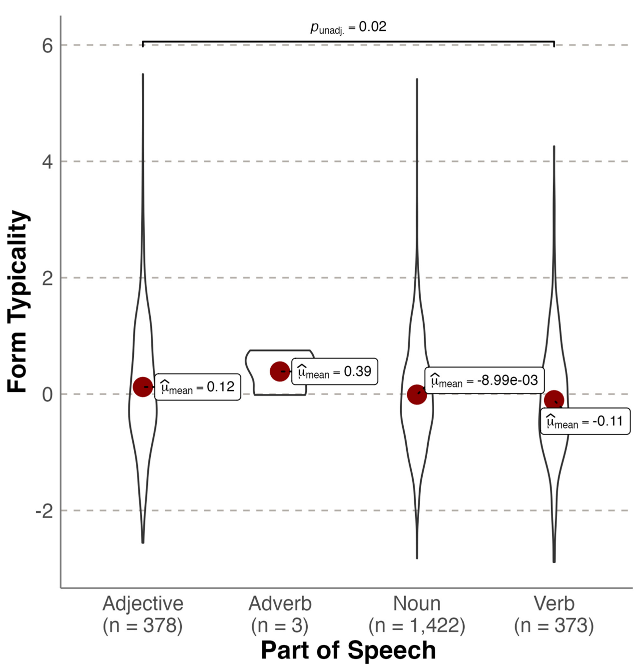
* $p < .05$. ** $p < .01$. *** $p < .001$.

were not significant (all *ps* > .05). Typicality for extremity of valence again differed significantly according to lexical category, Welch's $F(3, 9.98) = 11.039$, $p = .002$, est. $\omega^2 = .014$. Post hoc Games–Howell tests revealed adjectives were significantly more typical forms than nouns ($M_{\text{diff}} = 0.317$, $p < .001$) and verbs ($M_{\text{diff}} = 0.330$, $p < .001$). All other comparisons were not significant (all *ps* > .05).

The results of Study 1b indicate systematic form–meaning mappings account for 2% of variance in valence ratings of English monomorphemic words, a slightly smaller proportion to Study 1a,

Figure 4

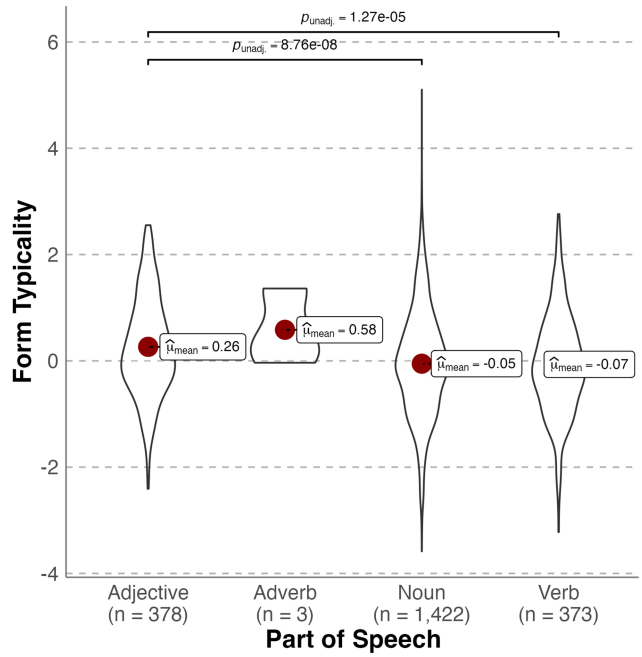
Violin Plots Showing Probability Densities of Valence Form Typicality Standardized Values as a Function of Lexical Category for Monomorphemic Words



Note. The shaded dot indicates the mean. p_{unadj} = unadjusted *p* value. See the online article for the color version of this figure.

Figure 5

Violin Plots Showing Probability Densities of Extremity of Valence Form Typicality Standardized Values as a Function of Lexical Category for Monomorphemic Words



Note. The shaded dot indicates the mean. p_{unadj} = unadjusted *p* value. See the online article for the color version of this figure.

and 1.3% in extremity of valence. Form variables explained less than half the variance in extremity of valence than in Study 1a, indicating morphophonological redundancy might have contributed to the former result. In addition, adjectives were on average more typical forms for both valence and extremity of valence. Notably, nouns and verbs did not differ in terms of their form typicality for both valence and extremity of valence, indicating form typicality for affective content differs to noun/verb (i.e., syntactic) form typicality in English (e.g., Sharpe & Marantz, 2017). Compared to Study 1a, the best-fit models also comprised fewer form variables. Table 5 shows the words with the top 10 most and least typical forms according to each measure.

We again replicated Louwerse and Qu's (2017) finding that nasals in the first phoneme predict negative valence. Hence, this finding cannot be attributed to morphophonological redundancies and does appear to be a distinctive feature of English words, at least for our sample comprising valence ratings with good agreement (cf., Adelman et al., 2018). Yet, nasals again predicted more neutral words in extremity of valence, suggesting a more complex relationship than mere sound symbolism. There was again no evidence of "front loading" of phonemes for valence as proposed by Adelman et al. (2018). We will return to this issue in the General Discussion section.

Study 2: Confirming Form-Valence Mappings of English Words With the NRC Valence, Arousal, and Dominance Lexicon Norms (S. Mohammad, 2018)

S. Mohammad (2018) recently provided the National Research Council Canada (NRC) valence, arousal, and dominance (VAD)

Table 5

Monomorphemic Words With the Top 10 Most (Positive Values) and Least (Negative Values) Typical Forms for Valence and Extremity of Valence

Valence				Extremity of valence			
Most typical forms		Least typical forms		Most typical forms		Least typical forms	
Word	Value	Word	Value	Word	Value	Word	Value
Easygoing	5.50	Describe	−2.89	Individuality	5.10	Match	−3.58
Goodwill	5.42	Neglect	−2.88	Vulnerability	3.48	Merge	−3.22
Videotape	4.63	Knob	−2.82	Helium	3.17	Stopwatch	−3.21
Emcee	4.32	Discard	−2.82	Deficiency	3.11	Patch	−3.02
Lightweight	4.30	Nod	−2.75	Enthusiasm	2.91	Perch	−3.02
Forthcoming	4.26	Newscaster	−2.68	Reinforce	2.76	Mortgage	−2.98
Be	4.26	Nutritious	−2.56	Insinuate	2.75	Magpie	−2.70
Widespread	3.82	Distinct	−2.54	Virtuoso	2.72	Page	−2.66
Homemade	3.51	Nest	−2.49	Heavyweight	2.67	Plunge	−2.66
Waive	3.38	Distinction	−2.41	Literacy	2.64	Sponge	−2.66

Lexicon comprising crowd-sourced emotional valence norms for 20,007 English words including those in the Warriner et al. (2013) data set. However, S. Mohammad (2018) used a different rating technique known as best-worse scaling (BWS) designed to address the limitations of traditional rating scales (Louivere et al., 2015). The BWS method involved presenting participants with sets of four words and having them select the two associated with the least and most happiness/pleasure/positiveness/satisfaction/contentedness/hopefulness, resulting in values of 0 (*negative*) and 1 (*positive*) indicative of opposite ends of the valence dimension. Despite differences in rating techniques, the two sets of valence norms are highly correlated ($r = .814$; see Table 4 in S. Mohammad, 2018). We, therefore, used the NRC VAD Lexicon valence norms to provide independent confirmation of the validity of our form typicality findings from Study 1b.[3] As lexical variables such as word length, age of acquisition, and frequency all correlate with valence (e.g., Kuperman et al., 2014), we also investigated whether the form typicality measures derived from the best-fit regression models in Study 1b are still able to predict both the Warriner et al. (2013) valence and extremity of valence values and the corresponding values from the NRC VAD norms after controlling for various sublexical and lexical variables (e.g., Adelman et al., 2018).

Method

Materials

The initial data set comprised 2,186 monomorphemic words from Study 1b. NRC VAD valence norms (S. Mohammad, 2018) were available for all of these words. The following sublexical and lexical variables were included as predictors: orthographic length (New et al., 2006), orthographic (OLD) and phonological Levenshtein (PLD) distances—the mean number of steps required through letter and phoneme substitutions, insertions, or transpositions to transform a word into its 20 closest neighbors (OLD20 and PLD20; Suárez et al., 2011; Yarkoni et al., 2008), Log10 average bigram frequency (Gao, Shinkareva, & Desai, 2022), SUBTLEX$_{US}$ frequency expressed as a Zipf score (Brysbaert & New, 2009; Zipf, 1949), prevalence—the number of people who know the word (Brysbaert et al., 2019), lexical category (part of speech classification; Brysbaert et al., 2012), age of acquisition (Kuperman et al., 2012), phonographic neighborhood size—the number of words differing in one letter and one phoneme from a target word (Adelman & Brown, 2007), feedforward (spelling-to-sound) and feedback (sound-to-spelling) word onset token consistency (Chee et al., 2020). All predictors were sourced from the South Carolina Psycholinguistic Database (SCOPE; Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022). This resulted in a final set of 2,127 words (370 adjectives, three adverbs, 1,389 nouns, and 365 verbs) including data from all predictor variables (Table 6). Extremity of valence from the NRC VAD norms was calculated by subtracting the midpoint of the scale (i.e., 0.5) from the valence rating of each word and disregarding polarity. Form typicality was calculated as the predicted value of the dependent variable for each word according to the relevant best-fit model in Study 1b (e.g., Valence Form Typicality = Valence + Residual; e.g., Sharpe & Marantz, 2017). The values for all words were then *Z*-transformed. Hence, positive typicality values indicate word forms aligned with positive valence and negative values indicate forms with negative valence. For extremity of valence ratings, positive typicality values indicate forms that are highly emotionally valenced regardless of polarity, while negative values indicate more neutral forms.

Design and Analysis

All analyses were performed in R Version 4.2.1 (R Core Team, 2022). We first calculated correlations between the respective valence and extremity of valence values from the two sets of norms. Next, we performed separate hierarchical linear regressions with robust *SE*s (Wilcox, 2016) with the valence and extremity of valence values from the Warriner et al. and NRC VAD norms as dependent variables with the packages *estimatr* (Blair et al., 2022) and *lmtest* (Zeileis & Hothorn, 2002). In each analysis, we entered the sublexical and lexical predictor variables in Step 1, followed by the form typicality measure.

Results and Discussion

Valence values from the Warriner et al. and NRC VAD (S. Mohammad, 2018) norms were strongly correlated ($r = .83$,

[3] The NRC VAD norms do not provide a measure of rating agreement/variability. Hence, we were not able to conduct analyses analogous to Study 1a with the valence and arousal ratings.

Table 6
Descriptive Statistics for the Variables in Study 2 (n = 2,127)

Variable	*M*	*SD*
Length	6.50	2.10
OLD	2.40	0.90
PLD	2.30	1.00
Log10 bigram frequency	2.80	0.42
Phonographic neighbors	2.10	3.90
Feedforward consistency	0.87	0.17
Feedback consistency	0.84	0.19
$\text{SUBTLEX}_{\text{US}}$ frequency	3.60	0.72
Age of acquisition	8.80	2.70
Prevalence	2.20	0.30
Valence (Warriner)	5.20	1.30
Valence (NRC)	0.52	0.21
Extremity of valence (Warriner)	1.00	0.85
Extremity of valence (NRC)	0.17	0.13
Form typicality (valence)	0.01	1.00
Form typicality (extremity of valence)	0.00	1.00
ELP latency (ms)	692	90
ELP accuracy	0.93	0.10

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ELP = English Lexicon Project.

$p < .001$), as were extremity of valence values ($r = .69$, $p < .001$) for the sample. The regression results for the form typicality measures with valence and extremity of valence are presented for the Warriner et al. norms in Tables 7 and 8, and for the NRC VAD norms in Tables 9 and 10, respectively.

Together, the sublexical and lexical variables significantly predicted valence and extremity of valence from the Warriner et al. (2013) norms, accounting for 7.1% and 8.2% of variance, respectively. However, of the individual variables, only length, feedforward spelling-sound consistency and age of acquisition were significant predictors for the former. For the latter, OLD, lexical frequency, prevalence, age of acquisition, and lexical category were all significant predictors. Form typicality explained 2.2% and 1.3% of unique variance in Warriner et al. (2013) valence and extremity of valence values, respectively.

Together, the sublexical and lexical variables significantly predicted valence and extremity of valence from the NRC VAD norms, accounting for 5.3% and 6.7% of variance, respectively. However, of the individual variables, only length, frequency, and age of acquisition were significant predictors for the former and only length and lexical frequency were significant predictors for the latter. Form typicality explained 1.7% and 1.2% of unique variance in NRC VAD valence and extremity of valence, respectively, despite the different rating methods used across normative studies (S. Mohammad, 2018; Warriner et al., 2013). Hence, the findings from Study 2 confirm the validity of the valence and extremity of valence form typicality measures from Study 1b.

Study 3: Written Lexical Decision

Study 3 investigated the extent to which valence and extremity of valence and their corresponding form typicality measures can predict written lexical decision performance from the ELP (Balota et al., 2007). Recent studies using the Warriner et al. (2013) norms that controlled for a range of sublexical and lexical variables in addition to arousal have shown negative words are responded to more slowly and/or that positive words are responded to more quickly (e.g., Barriga-Paulino et al., 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman et al., 2014; Scott et al., 2014) and that intensity of emotional valence regardless of polarity facilitates processing (e.g., Adelman & Estes, 2013; Kousta et al., 2009; Siakaluk et al., 2016; Vinson et al., 2014; Yap & Seow, 2014).[4] In addition, Kuperman et al. (2014; see also Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman, 2015) reported that valence and word frequency interact, resulting in stronger effects of valence among low- than high-frequency words. Due to the mixed findings above, and the questionable validity of the arousal ratings used in these studies (see Study 1a), we were agnostic about the direction of effects we would observe for valence using a hierarchical regression approach with the ELP data set. However, if both valence and extremity of valence and their corresponding measures of form typicality each contribute unique variance, we hypothesized that they would all be significant predictors of latencies and accuracy with the same set of words.

Method

Materials

The initial data set comprised 2,186 monomorphemic words from Study 1b. The following variables were included as controls: orthographic length (New et al., 2006), OLD20 and PLD20 (Suárez et al., 2011; Yarkoni et al., 2008), Log10 average bigram frequency (Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022), $\text{SUBTLEX}_{\text{US}}$ lexical frequency expressed as a Zipf score (Brysbaert & New, 2009; Zipf, 1949), Lexical Category (Brysbaert et al., 2012), phonographic neighborhood size (Adelman & Brown, 2007), feedforward (spelling-to-sound) and feedback (sound-to-spelling) word onset token consistency (Chee et al., 2020), age of acquisition (Kuperman et al., 2012), prevalence (Brysbaert et al., 2019), and concreteness—the degree to which the concept denoted by a word refers to a perceptible entity (Brysbaert et al., 2014). Valence and extremity of valence were taken from Warriner et al.'s (2013) norms. All predictors were sourced from the SCOPE database (Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022). Words in the ELP data set for which all the above variables were available were included in the study ($N = 2,094$; 364 adjectives, three adverbs, 1,365 nouns, and 362 verbs). Table 11 provides the descriptive statistics for each of the variables.

Design and Analysis

We performed separate hierarchical linear regressions with robust *SE*s (Wilcox, 2016) with two dependent variables from the ELP: latencies standardized as *z*-scores (*z*RT) and mean accuracy; with the packages *estimatr* (Blair et al., 2022) and *lmtest* (Zeileis &

[4] Previous studies using the ELP LDT data set included the quadratic term for valence in their linear regression models to investigate a nonlinear effect (e.g., Kuperman, 2015). Extremity (or intensity) of valence, while equivalent in shape to the quadratic term for valence, is a theoretically meaningful component of affective experience in its own right (see Kuppens et al., 2013; Murphy & Bastian, 2020). Hence, we opted to include it as a distinct variable in all our models.

Table 7
Regression Coefficients From Analysis of Warriner et al. (2013) Valence Norms (n = 2,127)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1				0.071***	
Length	0.084	0.032	2.630***		
OLD	−0.120	0.104	−1.157		
PLD	0.061	0.077	0.791		
Log10 bigram frequency	−0.013	0.088	−0.151		
Phonographic neighbors	−0.018	0.010	−1.845		
Feedforward consistency	−0.525	0.203	−2.580**		
Feedback consistency	−0.031	0.186	−0.168		
$\text{SUBTLEX}_{\text{US}}$ frequency	0.060	0.057	1.050		
Prevalence	0.001	0.099	0.014		
Age of acquisition	−0.134	0.014	−9.651***		
Lexical category	−0.020	0.036	−0.575		
Step 2				0.093***	0.022***
Typicality (valence)	0.203	0.028	7.268***		

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance.
** $p < .01$. *** $p < .001$.

Hothorn, 2002) in R Version 4.2.1 (R Core Team, 2022). In each analysis, we entered the control predictor variables in Step 1. Next, we entered valence and extremity of valence and their interactions with lexical frequency in Steps 2 and 3, followed by the corresponding measures of form typicality and their interactions with lexical frequency in Steps 4 and 5 (Model "a"). We then repeated these steps (Model "b"), reversing their order of entry (i.e., form typicality followed by valence measures). All predictor variables were mean-centered. Note that adopting this approach allows valence and extremity of valence and the measures of form typicality to each explain both their unique and shared variance with the other.

Results and Discussion

The results for latencies and accuracy for the ELP are presented in Tables 12 and 13, respectively.

Together the control predictor variables explained 58.4% and 51.3% of variance in ELP latencies and accuracy, respectively. Valence significantly predicted lexical decision latencies, explaining a small proportion (0.3%–4%) of variance, with more positive words being responded to more quickly (e.g., Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman, 2015; Kuperman et al., 2014). Figure 6 shows the relationship between response latencies and valence and extremity of valence form typicality after controlling for the sublexical and lexical predictor variables entered in Step 1. In addition, we observed a small but significant Valence × Frequency interaction consistent with prior work (0.1% of variance; e.g., Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman, 2015; Kuperman et al., 2014). While extremity of valence was a significant predictor when entered before the form typicality variables (0.1% of variance), it was no longer significant when entered after them. These results confirm previous findings indicating the effect of valence is best described by a monotonic rather than quadratic function (e.g., Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman et al., 2015). Only the form typicality for extremity of valence measure significantly predicted latencies (explaining 0.1% of variance) and showed a significant interaction with frequency (explaining 0.2% of variance). Valence, extremity of valence, and form typicality for valence did not significantly predict accuracy. Form typicality for

Table 8
Regression Coefficients From Analysis of Warriner et al. (2013) Extremity of Valence Norms (n = 2,127)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1				0.082***	
Length	−0.021	0.021	−0.975		
OLD	0.0129	0.065	1.984*		
PLD	0.016	0.047	0.346		
Log10 bigram frequency	−0.022	0.058	−0.381		
Phonographic neighbors	−0.006	0.006	−1.021		
Feedforward consistency	−0.058	0.128	−0.454		
Feedback consistency	−0.190	0.114	−1.671		
$\text{SUBTLEX}_{\text{US}}$ frequency	0.227	0.036	6.292***		
Prevalence	0.210	0.064	3.251**		
Age of acquisition	−0.023	0.009	−2.478*		
Lexical category	−0.109	0.020	−5.395***		
Step 2				0.095***	0.013***
Typicality (extremity of valence)	0.107	0.018	5.983***		

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 9

Regression Coefficients From Analysis of NRC VAD Valence Norms (n = 2,127)

Model comparison	Estimate	*SE*	*T*	Adjusted R^2	ΔR^2
Step 1				0.053***	
Length	0.017	0.005	3.289**		
OLD	−0.025	0.016	−1.526		
PLD	0.012	0.012	0.980		
Log10 bigram frequency	−0.006	0.013	−0.418		
Phonographic neighbors	−0.003	0.001	−1.882		
Feedforward consistency	−0.120	0.031	−3.872***		
Feedback consistency	−0.006	0.028	−0.196		
$\text{SUBTLEX}_{\text{US}}$ frequency	0.031	0.009	3.546***		
Prevalence	−0.005	0.017	−0.286		
Age of acquisition	−0.011	0.002	−5.008***		
Lexical category	−0.001	0.006	−0.115		
Step 2				0.070***	0.017***
Typicality (valence)	0.028	0.005	6.238***		

Note. VAD = valence, arousal, and dominance; OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance.
** $p < .01$. *** $p < .001$.

extremity of valence was a significant predictor (0.2% of variance). Valence showed a significant interaction with frequency, explaining 0.3%–0.5% of variance in accuracy.

Study 4: Written Word Recognition

The goal of Study 4 was essentially the same as that of Study 3 and used written word recognition performance from the ECP (Mandera et al., 2020), a behavioral megastudy in which participants were asked to indicate which words they knew. Mandera et al. showed that ELP and ECP response times are highly correlated despite participants not being asked to respond as quickly as possible, with ECP response times being on average 250 ms longer than their lexical decision equivalents. It is possible that this additional time could benefit processing of emotional valence and/or corresponding form information. Recently, Gao, Shinkareva, and Desai (2022) and Gao, Shinkareva, and Peelen (2022) reported that the valence effect for the ECP data set was not linear, unlike the ELP, with negative words showing a slower rate of change on response times than positive words. In addition, they did not observe an interaction between valence and lexical frequency. Our hypotheses were the same as in Study 3.

Method

Materials

The initial data set comprised 2,186 monomorphemic words from Study 1b. We included the same predictor variables as Study 3, excepting word prevalence (Brysbaert et al., 2019) because it is based on the same data set (see Mandera et al., 2020). This resulted in a data set of 2,093 words (Table 14; 363 adjectives, three adverbs, 1,365 nouns, and 355 verbs).

Design and Analysis

Identical to Study 3, except that the two dependent variables from the ECP were word recognition latencies standardized as *z*-scores (*z*RT) and mean accuracy.

Table 10

Regression Coefficients From Analysis of NRC VAD Extremity of Valence Norms (n = 2,127)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1				0.067***	
Length	0.007	0.003	2.162**		
OLD	0.000	0.010	0.042		
PLD	0.000	0.007	0.031		
Log10 bigram frequency	−0.006	0.008	−0.705		
Phonographic neighbors	−0.002	0.001	−1.818		
Feedforward consistency	−0.028	0.018	−1.553		
Feedback consistency	0.006	0.016	0.714		
$\text{SUBTLEX}_{\text{US}}$ frequency	0.038	0.005	7.249***		
Prevalence	0.020	0.010	1.883		
Age of acquisition	0.001	0.001	1.107		
Lexical category	−0.017	0.003	−5.405***		
Step 2				0.079***	0.012***
Typicality (extremity of valence)	0.015	0.003	5.411***		

Note. VAD = valence, arousal, and dominance; OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance.
** $p < .01$. *** $p < .001$.

Table 11
Descriptive Statistics for the Variables in Study 3 (n = 2,094)

Variable	*M*	*SD*
Length	6.50	2.10
OLD	2.40	0.90
PLD	2.30	1.00
Log10 bigram frequency	2.80	0.42
Phonographic neighbors	2.10	3.90
Feedforward consistency	0.87	0.17
Feedback consistency	0.84	0.19
$\text{SUBTLEX}_{\text{US}}$ frequency	3.60	0.72
Age of acquisition	8.70	2.70
Prevalence	2.20	0.28
Concreteness	3.50	1.10
Valence	5.20	1.30
Extremity of valence	1.00	0.85
Form typicality (valence)	0.00	1.00
Form typicality (extremity of valence)	0.00	1.00

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance.

Results and Discussion

The results for latencies and accuracy for the ECP are presented in Tables 15 and 16, respectively.

Together the control predictor variables explained 46.8% and 25.3% of variance in ECP latencies and accuracy, respectively. Both valence and extremity of valence significantly predicted latencies, explaining 0.6%–0.7% of variance in combination, with more positive and more extreme words being responded to more quickly (Figure 7). In addition, both variables showed significant interactions with frequency (0.9%–1.2% of variance). This pattern of results differs from that reported recently by Gao, Shinkareva, and Desai (2022) and Gao, Shinkareva, and Peelen (2022), who observed only a nonlinear effect of valence without any interactions with frequency. Form typicality for valence significantly predicted latencies when entered after the valence variables but not before (0.03%), while form typicality for extremity of valence showed a significant interaction with frequency (explaining 0.2%–0.4% of variance; Figure 7). Neither valence nor extremity of valence significantly predicted accuracy or significantly interacted with frequency.

Study 5a: Auditory Lexical Decision (AELP)

Our goal in Study 5a was to examine the extent to which valence, extremity of valence, and their corresponding measures of form typicality can predict auditory lexical decision performance from the AELP (Goh et al., 2020). Although the literature investigating the influence of emotional valence on spoken word recognition is smaller than the equivalent for written words, several studies have reported analogous findings. For example, Goh et al. (2016) reported both monotonic and quadratic patterns for the AELP database, Tucker et al. (2019) reported that positive valence was associated with quicker responses in auditory lexical decision (a monotonic function) using the MALD database, whereas Gao, Shinkareva, and Desai (2022) and Gao, Shinkareva, and Peelen (2022) reported that both negative and positive words facilitated performance (an inverted-U function commensurate with extremity

Table 12
Regression Coefficients From Item-Level Analyses of ELP LDT Latencies (n = 2,094)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.584***	
Step 2a (valence variables)				0.587***	0.003***
Val	−0.026	0.007	−3.608***		
ExtVal	0.023	0.012	1.978*		
Step 3a (interactions with frequency)				0.588***	0.001*
Val × Freq	0.024	0.009	2.732*		
ExtVal × Freq	−0.014	0.015	−0.960		
Step 4a (form typicality variables)				0.590***	0.002**
TypVal	0.008	0.010	0.841		
TypExtVal	0.030	0.010	2.975**		
Step 5a (interactions with frequency)				0.592***	0.002***
TypVal × Freq	0.005	0.013	0.347		
TypExtVal × Freq	−0.050	0.013	−3.896***		
Step 2b (form typicality variables)				0.585***	0.001**
TypVal	0.001	0.010	0.124		
TypExtVal	0.032	0.010	3.121**		
Step 3b (interactions with frequency)				0.587***	0.002***
TypVal × Freq	0.011	0.013	0.826		
TypExtVal × Freq	−0.051	0.013	−3.985***		
Step 4b (valence variables)				0.591***	0.004***
Val	−0.027	0.007	−3.719***		
ExtVal	0.021	0.012	1.767		
Step 5b (interactions with frequency)				0.592***	0.001*
Val × Freq	0.026	0.009	2.910**		
ExtVal × Freq	−0.005	0.015	−0.355		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; ELP = English Lexicon Project; LDT = lexical decision task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 13
Regression Coefficients From Item-Level Analyses of ELP LDT Accuracy (n = 2,094)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.511***	
Step 2a (valence variables)				0.511***	0.000
Val	0.000	0.001	0.797		
ExtVal	−0.002	0.002	−1.454		
Step 3a (interactions with frequency)				0.514***	0.003***
Val × Freq	−0.005	0.001	−4.198***		
ExtVal × Freq	−0.002	0.002	−0.995		
Step 4a (form typicality variables)				0.516***	0.002**
TypVal	−0.003	0.002	−1.660		
TypExtVal	−0.004	0.002	−2177*		
Step 5a (interactions with frequency)				0.517***	0.001
TypVal × Freq	0.004	0.002	1.594		
TypExtVal × Freq	0.002	0.002	0.745		
Step 2b (form typicality variables)				0.513***	0.002**
TypVal	−0.002	0.002	−1.394		
TypExtVal	−0.004	0.002	−2.328*		
Step 3b (interactions with frequency)				0.513***	0.000
TypVal × Freq	0.002	0.002	0.955		
TypExtVal × Freq	0.001	0.002	0.536		
Step 4b (valence variables)				0.512***	−0.001
Val	0.001	0.001	0.644		
ExtVal	−0.002	0.002	−1.147		
Step 5b (interactions with frequency)				0.517***	0.005***
Val × Freq	−0.006	0.001	−4.581***		
ExtVal × Freq	−0.003	0.002	−1.265		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; ELP = English Lexicon Project; LDT = lexical decision task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

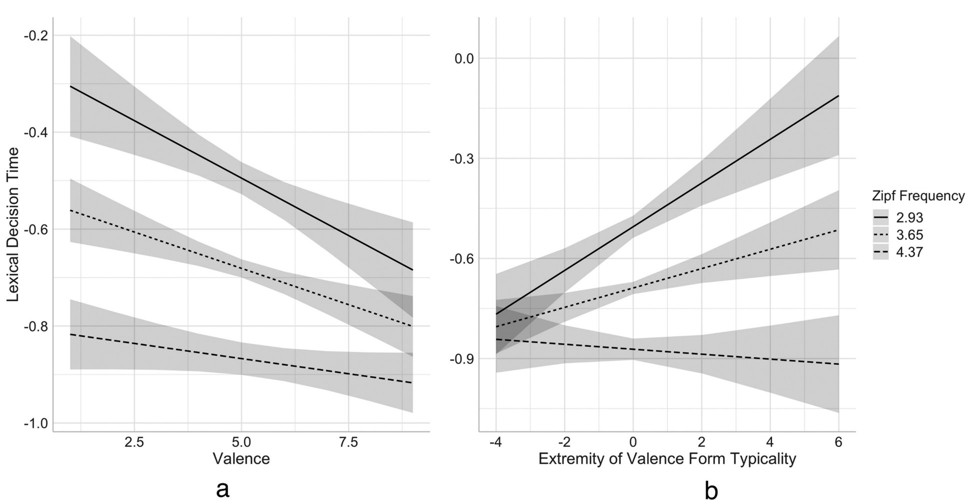
of valence) for both the MALD and AELP databases. Given the mixed findings, our hypotheses were essentially identical to Study 3. However, we expected form typicality would predict more variance in performance than in Study 3 given that auditory stimuli incorporate additional suprasegmental elements such as lexical stress.

Method

Materials

The initial data set comprised 2,186 monomorphemic words from Study 1b. Predictor variables were identical to Study 3 with the addition of American Dialect auditory token duration. Words in the

Figure 6
Added Variable Plot Showing the Relationships Between ELP Lexical Decision Latencies and (a) Valence and (b) Extremity of Valence Form Typicality as a Function of Lexical Frequency After Controlling for the Sublexical and Lexical Predictor Variables



Note. Shaded area shows 95% confidence intervals. ELP = English Lexicon Project.

Table 14

Descriptive Statistics for the Variables in Study 4 (n = 2,093)

Variable	*M*	*SD*
Length	6.50	2.10
OLD	2.40	0.90
PLD	2.30	1.00
Log10 bigram frequency	2.80	0.42
Phonographic neighbors	2.10	3.90
Feedforward consistency	0.87	0.17
Feedback consistency	0.84	0.19
$\text{SUBTLEX}_{\text{US}}$ frequency	3.60	0.72
Age of acquisition	8.70	2.70
Concreteness	3.50	1.10
Valence	5.20	1.30
Extremity of valence	1.00	0.85
Form typicality (valence)	0.00	1.00
Form typicality (extremity of valence)	0.00	1.00
ECP latency (milliseconds)	904	85
ECP accuracy	0.98	0.02

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ECP = English Crowdsourcing Project.

AELP data set for which the variables were available were included in the study (N = 1,586; 264 adjectives, three adverbs, 1,021 nouns, and 298 verbs). Table 17 provides the descriptive statistics for each of the variables.

Design and Analysis

Identical to Study 3, except that the two dependent variables from the AELP were American Dialect latencies standardized as z-scores (zRT) and mean accuracy.

Results and Discussion

The results for latencies and accuracy for the AELP are presented in Tables 18 and 19, respectively.

Together the control predictor variables explained 22.6% and 21.9% of variance in AELP latencies and accuracy, respectively. Valence did not significantly predict latencies although extremity of valence did, explaining an additional 0.6%–0.9% of variance (Figure 8a), consistent with both Goh et al.'s (2016) and Gao, Shinkareva, and Desai's (2022) and Gao, Shinkareva, and Peelen's (2022) findings that both negative and positive valence facilitated latencies compared to neutral words. Valence significantly interacted with frequency only after being entered following the form typicality variables (0.2% of variance). Form typicality for valence significantly predicted latencies, with more typical forms facilitating responses (1.7%–1.9% of variance; Figure 8b), whereas typicality for extremity of valence did not. Only form typicality significantly interacted with frequency when entered after the valence variables. Neither of the valence variables were significant predictors of AELP accuracy or interacted significantly with frequency. However, form typicality for extremity of valence significantly interacted with frequency, predicting an additional 0.3% of variance.

Study 5b: Auditory Lexical Decision (MALD)

Study 5b was essentially a repeat of Study 5a with auditory lexical decision performance from the MALD database (Tucker et al., 2019). Hypotheses were therefore identical to Study 5a.

Table 15

Regression Coefficients From Item-Level Analyses of ECP Recognition Latencies (n = 2,093)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.468***	
Step 2a (valence variables)				0.474***	0.006***
Val	−0.006	0.001	−4.413***		
ExtVal	−0.008	0.002	−3.563***		
Step 3a (interactions with frequency)				0.483***	0.009***
Val × Freq	0.009	0.002	4.974***		
ExtVal × Freq	0.011	0.003	3.733***		
Step 4a (form typicality variables)				0.486***	0.003**
TypVal	0.005	0.002	1.954		
TypExtVal	0.004	0.002	1.944		
Step 5a (interactions with frequency)				0.490***	0.004***
TypVal × Freq	−0.004	0.003	−1.235		
TypExtVal × Freq	−0.009	0.003	−3.237**		
Step 2b (form typicality variables)				0.469***	0.001
TypVal	0.003	0.002	1.159		
TypExtVal	0.003	0.002	1.596		
Step 3b (interactions with frequency)				0.471***	0.002**
TypVal × Freq	−0.004	0.004	−0.634		
TypExtVal × Freq	−0.008	0.003	−2.721**		
Step 4b (valence variables)				0.478***	0.007***
Val	−0.006	0.001	−4.671***		
ExtVal	−0.008	0.002	−3.734***		
Step 5b (interactions with frequency)				0.490***	0.012***
Val × Freq	0.010	0.002	5.415***		
ExtVal × Freq	0.012	0.003	4.353***		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; ECP = English Crowdsourcing Project.
** $p < .01$. *** $p < .001$.

Table 16

Regression Coefficients From Item-Level Analyses of ECP Recognition Accuracy (n = 2,093)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.253***	
Step 2a (valence variables)				0.254***	0.001
Val	0.000	0.000	0.594		
ExtVal	0.001	0.000	2.409*		
Step 3a (interactions with frequency)				0.271***	0.017***
Val × Freq	−0.002	0.000	−4.903***		
ExtVal × Freq	−0.004	0.001	−5.952***		
Step 4a (form typicality variables)				0.276***	0.005
TypVal	−0.002	0.001	−1.952		
TypExtVal	−0.000	0.001	−0.792		
Step 5a (interactions with frequency)				0.282***	0.006*
TypVal × Freq	0.003	0.001	1.925		
TypExtVal × Freq	0.001	0.001	1.197		
Step 2b (form typicality variables)				0.258***	0.005
TypVal	−0.002	0.001	−1.853		
TypExtVal	0.001	0.001	−0.636		
Step 3b (interactions with frequency)				0.261***	0.003
TypVal × Freq	0.002	0.001	1.589		
TypExtVal × Freq	0.000	0.001	0.356		
Step 4b (valence variables)				0.262***	0.001*
Val	0.000	0.000	0.180		
ExtVal	0.001	0.000	2.482*		
Step 5b (interactions with frequency)				0.282***	0.020***
Val × Freq	−0.002	0.000	−5.297***		
ExtVal × Freq	−0.004	0.001	−6.257***		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; ECP = English Crowdsourcing Project.
* $p < .05$. *** $p < .001$.

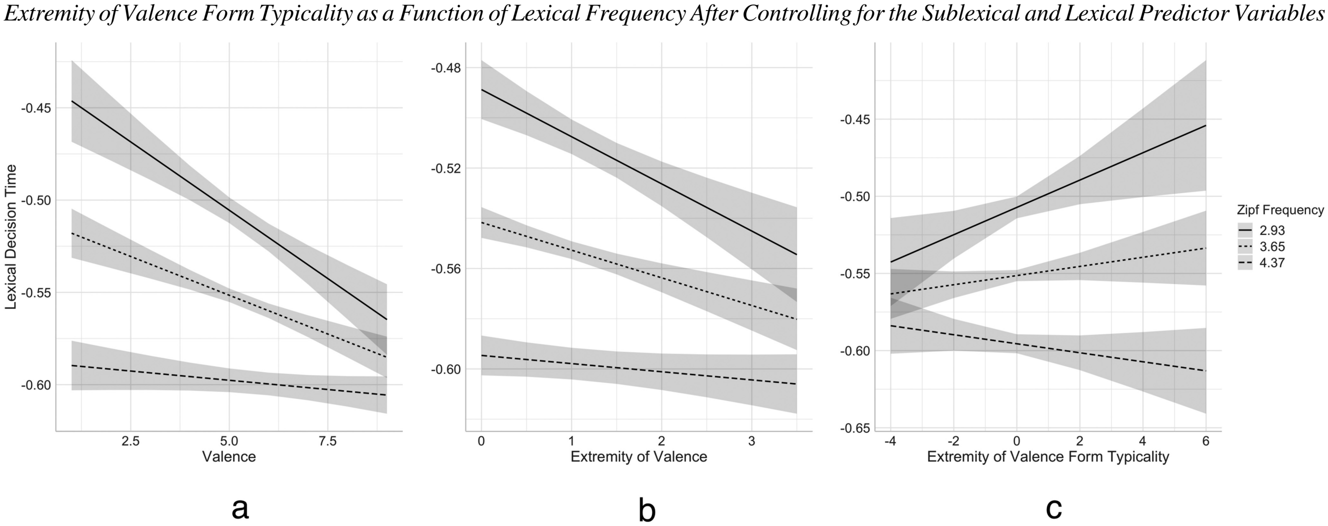
Method

Materials

The initial data set comprised 2,186 monomorphemic words from Study 1b. All predictor variables were identical to Study 5a. Words in the MALD data set for which the variables were available were included in the study ($N =$ 2,092; 363 adjectives, three adverbs, 1,365 nouns, and 361 verbs). Table 20 provides the descriptive statistics for each of the variables.

Figure 7

Added Variable Plot Showing the Relationship Between ECP Word Recognition Latencies and (a) Valence, (b) Extremity of Valence, and (c) Extremity of Valence Form Typicality as a Function of Lexical Frequency After Controlling for the Sublexical and Lexical Predictor Variables



Note. Shaded area shows 95% confidence intervals. ECP = English Crowdsourcing Project.

Table 17
Descriptive Statistics for the Variables in Study 5a (n = 1,586)

Variable	*M*	*SD*
Length	6.30	2.00
Duration (ms)	786	129
OLD	2.30	0.84
PLD	2.20	0.97
Log10 bigram frequency	2.80	0.40
Phonographic neighbors	2.30	4.10
Feedforward consistency	0.87	0.17
Feedback consistency	0.85	0.18
$\text{SUBTLEX}_{\text{US}}$ frequency	3.80	0.72
Age of acquisition	8.20	2.50
Prevalence	2.30	0.21
Concreteness	3.50	1.10
Valence	5.20	1.40
Extremity of valence	1.10	0.87
Form typicality (valence)	−0.01	0.97
Form typicality (extremity of valence)	−0.06	0.99
AELP latency (ms)	967	91
AELP accuracy	0.93	0.10

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; AELP = Auditory English Lexicon Project.

Design and Analysis

Identical to Study 5a, except that the two dependent variables from the MALD were latencies standardized as *z*-scores (*z*RT) and mean accuracy.

Results and Discussion

The results for latencies and accuracy for the MALD are presented in Tables 21 and 22, respectively.

Together the control predictor variables explained 7.5% and 11.2% of variance in MALD latencies and accuracy, respectively. Both valence and extremity of valence significantly predicted latencies and accuracy, explaining an additional 0.3%–0.5% of variance in combination, consistent with Tucker et al.'s (2019) and Gao, Shinkareva, and Desai's (2022) and Gao, Shinkareva, and Peelen's (2022) findings with the MALD database. More positive and emotional words facilitated latencies compared to neutral words (Figure 9). Neither significantly interacted with frequency. Form typicality for valence significantly predicted latencies, with more typical forms facilitating responses (0.9%–1.1% of variance), whereas typicality for extremity of valence did not. Neither of the form typicality measures significantly interacted with frequency. None of the valence or form typicality variables were significant predictors of MALD accuracy or interacted significantly with frequency.

Study 6: Reading Aloud

Our goal in Study 6 was to examine the extent to which valence, extremity of valence and their corresponding form typicality measures can predict reading-aloud performance (i.e., word naming/pronunciation) from the ELP (Balota et al., 2007). Algom et al. (2004) reported that negative words were articulated more slowly than neutral words, and Estes and Adelman (2008a, 2008b) reported negative words were articulated more slowly than positive words, while

Table 18
Regression Coefficients From Item-Level Analyses of AELP LDT Latencies (n = 1,586)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.226***	
Step 2a (valence variables)				0.235***	0.009***
Val	−0.005	0.006	−0.927		
ExtVal	−0.043	0.010	−4.483***		
Step 3a (interactions with frequency)				0.237***	0.002*
Val × Freq	0.13	0.007	1.827		
ExtVal × Freq	−0.020	0.012	−1.645		
Step 4a (form typicality variables)				0.254***	0.017***
TypVal	−0.048	0.009	−5.265***		
TypExtVal	0.009	0.008	−1.051		
Step 5a (interactions with frequency)				0.257***	0.003
TypVal × Freq	−0.029	0.014	−2.055*		
TypExtVal × Freq	0.004	0.012	0.306		
Step 2b (form typicality variables)				0.245***	0.019***
TypVal	−0.049	0.009	−5.387***		
TypExtVal	−0.012	0.008	1.431		
Step 3b (interactions with frequency)				0.247***	0.002
TypVal × Freq	−0.026	0.014	−1.860		
TypExtVal × Freq	−0.001	0.012	−0.091		
Step 4b (valence variables)				0.254***	0.006***
Val	0.002	0.006	0.267		
ExtVal	−0.038	0.010	−3.985***		
Step 5b (interactions with frequency)				0.256***	0.001*
Val × Freq	0.015	0.007	2.043***		
ExtVal × Freq	−0.019	0.012	−1.494		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; AELP = Auditory English Lexicon Project; LDT = lexical decision task.
* $p < .05$. *** $p < .001$.

Table 19

Regression Coefficients From Item-Level Analyses of AELP LDT Accuracy (n = 1,586)

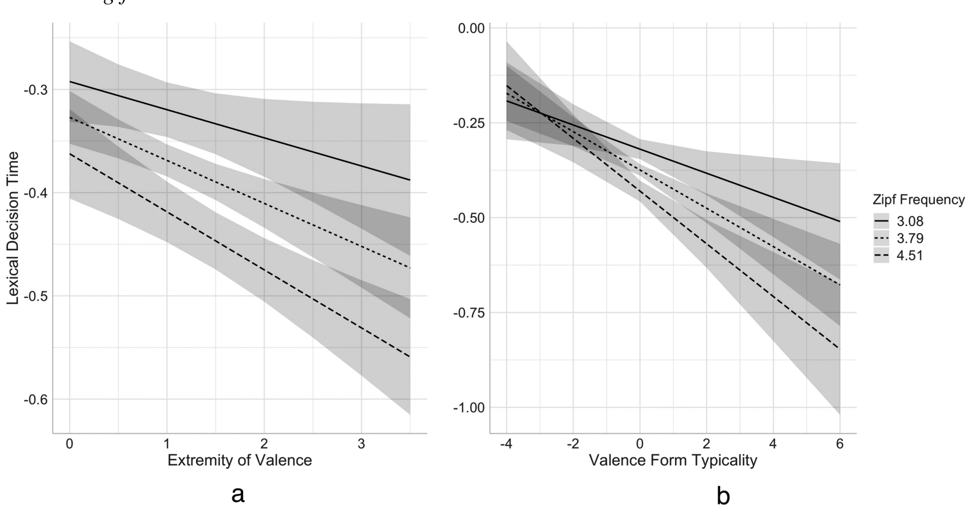
Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.219***	
Step 2a (valence variables)				0.219***	0.000
Val	0.000	0.001	−0.239		
ExtVal	0.004	0.002	1.615		
Step 3a (interactions with Frequency)				0.219***	0.000
Val × Freq	−0.002	0.002	−1.056		
ExtVal × Freq	−0.004	0.003	−1.172		
Step 4a (form typicality variables)				0.219***	0.000
TypVal	−0.004	0.002	−1.521		
TypExtVal	0.003	0.002	1.038		
Step 5a (interactions with frequency)				0.222***	0.003*
TypVal × Freq	0.006	0.004	1.601		
TypExtVal × Freq	−0.008	0.003	−2.724**		
Step 2b (form typicality variables)				0.219***	0.000
TypVal	−0.004	0.002	−1.484		
TypExtVal	0.003	0.002	1.168		
Step 3b (interactions with frequency)				0.222***	0.003**
TypVal × Freq	0.005	0.003	1.452		
TypExtVal × Freq	−0.008	0.003	−2.845**		
Step 4b (valence variables)				0.222***	0.000
Val	0.000	0.002	0.049		
ExtVal	0.004	0.002	1.682		
Step 5b (interactions with frequency)				0.222***	0.000
Val × Freq	−0.003	0.002	−1.300		
ExtVal × Freq	−0.003	0.003	−0.831		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; AELP = Auditory English Lexicon Project; LDT = lexical decision task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Kuperman et al. (2014) reported positive valence facilitated pronunciation of words in the ELP data set. To the best of our knowledge, extremity of valence has yet to be investigated in relation to reading aloud. As per our previous studies of word recognition, we hypothesized that valence, extremity of valence and their corresponding measures of form typicality would be significant predictors of latencies and accuracy with the same set of words if they each contributed unique variance.

Figure 8

Added Variable Plot Showing the Relationship Between AELP Lexical Decision Latencies and (a) Extremity of Valence and (b) Form Typicality for Valence as a Function of Lexical Frequency After Controlling for the Sublexical and Lexical Predictor Variables



Note. Shaded area shows 95% confidence intervals. AELP = Auditory English Lexicon Project.

Table 20
Descriptive Statistics for the Variables in Study 5b (n = 2,092)

Variable	*M*	*SD*
Length	6.50	2.10
Duration (ms)	513	122
OLD	2.40	0.90
PLD	2.30	1.00
Log10 bigram frequency	2.80	0.42
Phonographic neighbors	2.10	3.90
Feedforward consistency	0.87	0.17
Feedback consistency	0.84	0.19
$\text{SUBTLEX}_{\text{US}}$ frequency	3.60	0.72
Age of acquisition	8.70	2.70
Prevalence	2.20	0.28
Concreteness	3.50	1.10
Valence	5.20	1.30
Extremity of valence	1.00	0.85
Form typicality (valence)	0.00	1.00
Form typicality (extremity of valence)	0.00	1.00
MALD latency (ms)	897	147
MALD accuracy	0.95	0.12

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; MALD = Massive Auditory Lexical Decision.

Method

Materials

The data set comprised the same set of 2,094 words from Study 3. Predictor variables were identical to Study 3. Mean ELP naming latency was 661 ms (*SD* = 73) and accuracy 0.98 (*SD* = 0.05).

Design and Analysis

Identical to Study 3, except that the two dependent variables from the ELP were for reading aloud: latencies standardized as *z*-scores (*z*RT) and mean accuracy.

Results and Discussion

The results for latencies and accuracy are presented in Tables 23 and 24, respectively.

Together the control predictor variables explained 47.6% and 27.2% of variance in ELP reading aloud latencies and accuracy, respectively. Surprisingly, neither valence nor extremity of valence significantly predicted latencies or accuracy (cf., Algom et al., 2004; Estes & Adelman, 2008a, 2008b; Kuperman et al., 2014). In addition, neither interacted with frequency to significantly predict additional variance. However, form typicality for valence significantly predicted both, with more positive forms being produced more quickly yet less accurately (Figure 10), explaining 0.2% and 0.4% of variance, respectively. While typicality for extremity of valence did not significantly predict either latencies or accuracy, it did interact significantly with frequency for latencies, explaining an additional 0.3% of variance.

Study 7: Recognition Memory

Some theoretical accounts of memory explicitly propose an advantage for negative words (e.g., Bowen et al., 2018). However, the empirical literature on emotion and memory has been inconsistent at best (for a review, see MacMillan et al., 2022). Adelman and Estes (2013) reported a recognition memory advantage for extremity of valence that has recently been replicated by two studies using megastudy data sets (Cortese & Khanna, 2022; Lau et al., 2018). As per our previous studies, we hypothesized that valence, extremity of valence, and their corresponding form typicality measures would be significant predictors of performance with the same set of words if they each contributed unique variance.

Method

Materials

The initial data set comprised 2,186 monomorphemic words from Study 1b. Cortese and Khanna's (2022) recognition memory megastudy database comprises 5,587 words from two separate experiments employing monosyllabic or disyllabic words (Cortese et al., 2010, 2015). In each experiment, 120 participants were required to learn 30 study lists of 50 words, with each written word presented individually for a brief period. They were then tested on 30 lists of 100 words. Test lists comprised 1,500 study list words plus another 1,500 unstudied words (i.e., another 30 lists of 50 words), with participants required to make an "old" (i.e., studied) or "new" (i.e., unstudied) decision for each word. Between each study and test list, participants confirmed the accuracy of 18 simple maths problems. A brief rest break occurred between study-test lists. Both experiments were conducted over 2 days (15 lists per day). Predictor variables were identical to Studies 3, 5, and 6. Words in Cortese and Khanna's (2022) megastudy database for which these variables were available were included in the study (total *N* = 1,211; 148 adjectives, three adverbs, 788 nouns, 272 verbs). Table 25 provides the descriptive statistics for each of the variables in the study.

Design and Analysis

Identical to Studies 3 and 5, except that the dependent variable was corrected hit probability (hits minus false alarms).

Results and Discussion

The results are presented in Table 26. Combined, the control variables predicted 27.8% of the variance in corrected hit probability. Of the two valence variables, only extremity of valence significantly predicted an additional 3%–3.3% of variance, with more emotional words being better remembered (see Figure 11), replicating prior work (e.g., Adelman & Estes, 2013; Cortese & Khanna, 2022; MacMillan et al., 2022). Neither valence variable interacted significantly with frequency. Form typicality for valence also predicted memory performance significantly (0.9%–1.3% of variance). More positive forms were better remembered (Figure 11). However, typicality for extremity of valence did not contribute significantly. There were no significant interactions with frequency for either form typicality variable.

General Discussion

It is generally accepted that a word's emotional valence influences how it is processed and remembered, yet the evidence is weak and often contradictory. The primary aim of the present paper was to determine whether surface form variables are able to predict the

Table 21
Regression Coefficients From Item-Level Analyses of MALD Project LDT Latencies ($n = 2{,}092$)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.075***	
Step 2a (valence variables)				0.080***	0.005**
Val	−0.028	0.013	−2.195*		
ExtVal	−0.066	0.020	−3.265**		
Step 3a (interactions with frequency)				0.079***	−0.001
Val × Freq	0.005	0.017	0.317		
ExtVal × Freq	−0.005	0.027	0.182		
Step 4a (form typicality variables)				0.088***	0.009***
TypVal	−0.080	0.017	−4.572***		
TypExtVal	−0.005	0.018	0.294		
Step 5a (interactions with frequency)				0.087***	−0.001
TypVal × Freq	−0.005	0.023	−0.216		
TypExtVal × Freq	0.014	0.023	0.636		
Step 2b (form typicality variables)				0.086***	0.011***
TypVal	−0.084	0.017	−4.851***		
TypExtVal	−0.011	0.018	−0.593		
Step 3b (interactions with frequency)				0.085***	−0.001
TypVal × Freq	−0.007	0.023	−0.331		
TypExtVal × Freq	0.010	0.022	0.438		
Step 4b (valence variables)				0.088***	0.003**
Val	−0.018	0.013	−1.382		
ExtVal	−0.061	0.020	−2.980**		
Step 5b (interactions with frequency)				0.087***	−0.001
Val × Freq	0.002	0.018	0.098		
ExtVal × Freq	−0.007	0.027	0.271		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; MALD = Massive Auditory Lexical Decision; LDT = lexical decision task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

rated emotional valence of English words (Warriner et al., 2013). We found evidence of significant nonarbitrary statistical relationships between form features and both valence and extremity of valence that we confirmed with an independent set of similar ratings (S. Mohammad, 2018). Additionally, we demonstrated for the first time that form typicality for valence and extremity of valence were significant predictors of lexical access during word comprehension and production as well as recognition memory across several behavioral megastudy data sets.

Prior work on emotional sound symbolism reported nonarbitrary relationships between phonemes and primarily negative valence (e.g., Adelman et al., 2018; Louwerse & Qu, 2017; but see Yu et al., 2021). In Studies 1 and 2, we found that surface form variables predict approximately 2% of English monomorphemic words' rated valence across two sets of independent norms (S. Mohammad, 2018; Warriner et al., 2013), even when other variables that correlate with valence, such as lexical frequency and age of acquisition, are controlled. Specifically, we found that positive words tend to comprise more syllables, and more bilabial and velar sounds in the initial phoneme, as well as more labiodental sounds in the final phoneme, whereas negative words are more likely to have a syllable that is stressed, comprise more stops and fricatives, and be more likely to have nasal sounds in the initial phoneme.

We were unable to replicate Adelman et al.'s (2018) finding of "front loading" that they proposed to be a relatively distinctive property of emotional sound symbolism in English. We could also find no evidence for high or low vowels being able to predict valence (cf., Yu et al., 2021). However, we did replicate Louwerse and Qu's (2017) finding that nasals in the first phoneme predict negative valence in English, as well as their result for velars in the first phoneme that predict positive valence. Furthermore, Study 1b showed the result for nasals cannot be attributed to morphophonological redundancies in negating affixes ("in-," "un-," and "im-") in English as Adelman et al. (2018) proposed. The most likely reason for the discrepancy in findings across studies is the use of words whose rating *SD* indicated poor agreement about their emotional valence and arousal properties to begin with (Pollock, 2018). Louwerse and Qu (2017) included only words with extreme valence values that are more likely to have lower *SD*s, whereas Adelman and Estes (2013) employed most of the words in the Warriner et al. (2013) norms and included arousal ratings in their regression models.

We also investigated whether surface form variables could predict extremity of valence, that is emotional strength regardless of polarity (e.g., Adelman & Estes, 2013). Here too we found evidence of significant form typicality, with form variables explaining 1.3% of variance for ratings of monomorphemic words. Emotional words tended to have more syllables and were more likely to have voiced sounds in their initial phoneme, whereas neutral words were more likely to comprise nasal sounds in their initial phoneme. The latter result is interesting given nasals in the first phoneme were also a significant predictor of negative valence along with other features such as the application of lexical stress (e.g., Louwerse & Qu, 2017). This suggests a more complex relationship than mere sound symbolism. The combined results for valence and extremity of valence suggest nasals in the initial phoneme might be more likely to signal neutral emotion unless the word also includes stress on that phoneme and/or additional features. Interestingly, adjectives were the most form-typical

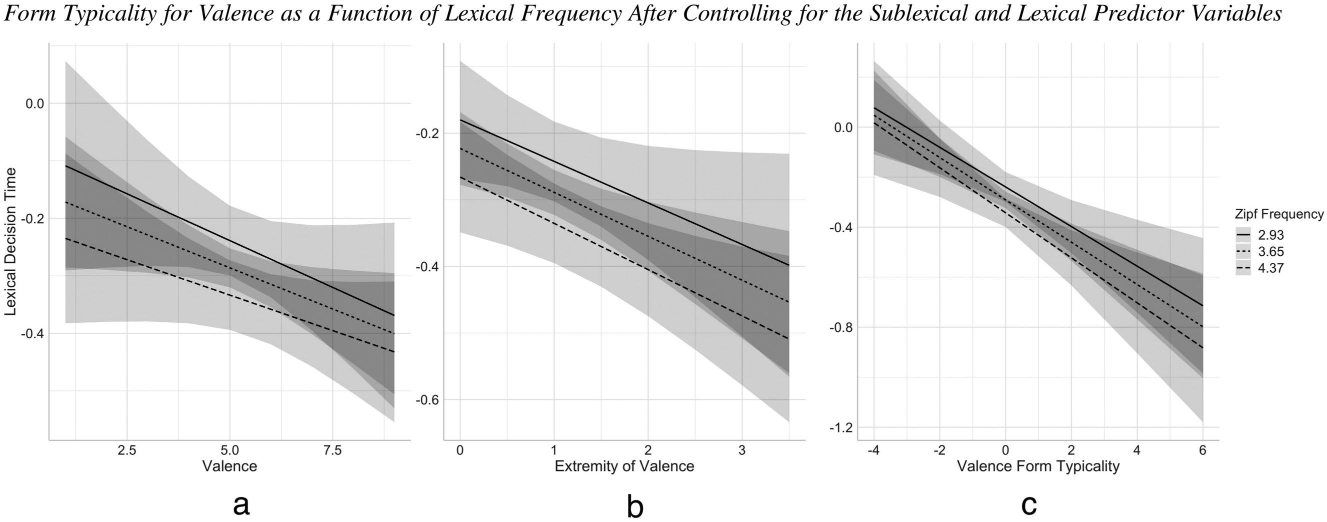
Table 22
Regression Coefficients From Item-Level Analyses of MALD Project LDT Accuracy ($n = 2,092$)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.112***	
Step 2a (valence variables)				0.113***	
Val	0.000	0.002	−0.229		
ExtVal	0.006	0.003	2.139*		
Step 3a (interactions with frequency)				0.112***	−0.001
Val × Freq	−0.002	0.002	−0.764		
ExtVal × Freq	−0.003	0.003	−0.968		
Step 4a (form typicality variables)				0.112***	0.000
TypVal	−0.001	0.003	−0.507		
TypExtVal	0.001	0.003	0.408		
Step 5a (interactions with frequency)				0.112***	0.000
TypVal × Freq	0.006	0.003	1.961		
TypExtVal × Freq	−0.004	0.004	−1.076		
Step 2b (form typicality variables)				0.111***	−0.001
TypVal	−0.001	0.003	−0.501		
TypExtVal	0.002	0.003	0.587		
Step 3b (interactions with frequency)				0.112***	0.000
TypVal × Freq	0.006	0.003	1.910		
TypExtVal × Freq	−0.004	0.004	−1.160		
Step 4b (valence variables)				0.112***	0.000
Val	0.000	0.002	−0.154		
ExtVal	0.006	0.003	2.102*		
Step 5b (interactions with frequency)				0.112***	0.000
Val × Freq	−0.002	0.002	−1.028		
ExtVal × Freq	−0.003	0.003	−0.838		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency; MALD = Massive Auditory Lexical Decision; LDT = lexical decision task.
* $p < .05$. *** $p < .001$.

lexical category, which is perhaps not surprising given their roles in directly referencing emotional states and feelings (e.g., Béligon, 2020). However, the overall proportion of variance explained by surface form variables (∼2%) was quite modest compared to those reported for noun/verb status (12.4%; Sharpe & Marantz, 2017), sensory experience (12.3%; G. de Zubicaray, Arciuli, Guenther, et al., 2023) and semantic size ratings (20%; G. I. de Zubicaray, Kearney, Arciuli, et al., 2023). This might be because suprasegmental features can be used more flexibly to convey information about emotion through affective prosody independent of

Figure 9
Added Variable Plot Showing the Relationship Between MALD Lexical Decision Latencies and (a) Valence, (b) Extremity of Valence, and (c) Form Typicality for Valence as a Function of Lexical Frequency After Controlling for the Sublexical and Lexical Predictor Variables



Note. Shaded area shows 95% confidence intervals. MALD = Massive Auditory Lexical Decision.

Table 23
Regression Coefficients From Item-Level Analyses of ELP Reading Aloud Latencies (n = 2,094)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.476***	
Step 2a (valence variables)				0.477***	0.001
Val	−0.015	0.008	−1.897		
ExtVal	−0.015	0.013	−1.183		
Step 3a (interactions with frequency)				0.478***	0.001
Val × Freq	0.019	0.010	1.810		
ExtVal × Freq	−0.019	0.016	−1.179		
Step 4a (form typicality variables)				0.478***	0.000
TypVal	−0.026	0.012	−2.205*		
TypExtVal	0.010	0.012	0.862		
Step 5a (interactions with frequency)				0.481***	0.003***
TypVal × Freq	−0.009	0.018	−0.536		
TypExtVal × Freq	−0.054	0.015	−3.653***		
Step 2b (form typicality variables)				0.478***	0.002*
TypVal	−0.030	0.012	−2.545*		
TypExtVal	0.009	0.012	0.780		
Step 3b (interactions with frequency)				0.481***	0.003***
TypVal × Freq	−0.005	0.017	−0.308		
TypExtVal × Freq	−0.056	0.015	−3.839***		
Step 4b (valence variables)				0.481***	0.000
Val	−0.011	0.008	−1.342		
ExtVal	−0.012	0.013	−0.928		
Step 5b (interactions with frequency)				0.481***	0.000
Val × Freq	0.020	0.010	1.985*		
ExtVal × Freq	−0.009	0.017	−0.551		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency, ELP = English Lexicon Project.
* $p < .05$. *** $p < .001$.

word meaning (e.g., Frick, 1985; Kamiloğlu et al., 2020; S. K. Kim & Sumner, 2017).

The findings from Study 3 with the written LDT were consistent with the subset of studies reviewed in the Introduction that reported a linear function of valence, with more positive words responded to more quickly and more negative words responded to more slowly than neutral words. However, the proportion of variance explained was quite small (0.3%). When expressed as a proportion of the mean ELP response latency in Study 3 (692 ms), the effect size for valence translates to approximately 2 ms of processing time. In addition, we observed a small but significant interaction with lexical frequency such that the effect of valence was greater for low-frequency words, consistent with previous reports (e.g., Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022; Kuperman et al., 2014). Only form typicality for extremity of valence significantly predicted latencies and accuracy (explaining 0.1% and 0.2% of variance, respectively), and showed a significant interaction with frequency (explaining 0.2% of variance). Study 4 with the ECP data set revealed a different pattern of results, with both valence and extremity of valence significantly predicting written word recognition latencies and interacting with frequency. These findings differ to those of Gao, Shinkareva, and Desai (2022) and Gao, Shinkareva, and Peelen (2022) who recently reported only a nonlinear effect of valence on ECP latencies. The discrepancy in results is likely attributable to the latter study's use of a larger sample of mono- and multimorphemic words, including many with poor rating agreement, and inclusion of arousal ratings in their model. Form typicality for either emotion variable did not emerge as a significant predictor for ECP performance, although form typicality for extremity of valence did interact significantly with frequency, mirroring the ELP results. Overall, the results indicate that valence, extremity of valance and their phonological form typicality counterparts collectively have relatively little influence on visual lexical decision and word recognition performance. As Mandera et al. (2020) noted, responses on these tasks can be based primarily on orthographic activation.

Across Studies 5 and 5b, extremity of valence was a significant predictor of auditory lexical decision latencies in both the AELP and MALD data sets, respectively, explaining a small proportion of variance (0.5%–0.9%), consistent with Goh et al.'s (2016) and Gao, Shinkareva, and Desai's (2022) and Gao, Shinkareva, and Peelen's (2022) findings (cf. Tucker et al., 2019). In Study 5b, valence also emerged as a significant predictor of MALD latencies consistent with Tucker et al. (2019; cf. Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022). As Gao, Shinkareva, and Desai (2022) and Gao, Shinkareva, and Peelen (2022) noted, variations in results across auditory LDT studies might be due to differences in the set of words studied and selection of control variables. The present study differed to the previous auditory LDT studies of valence effects by including only monomorphemic words with good rating agreement (Pollock, 2018) and word prevalence as a lexical control variable (Brysbaert et al., 2019), with the latter able to explain a significant proportion of variance across all our behavioral megastudy data sets prior to entering valence in the model. Only form typicality for extremity of valence was able to significantly predict auditory LDT latencies, explaining a relatively larger but still small proportion of variance (1.9% and 1.1% for AELP and MALD,

Table 24

Regression Coefficients From Item-Level Analyses of ELP Reading Aloud Accuracy (n = 2,094)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.272***	
Step 2a (valence variables)				0.272***	0.000
Val	−0.001	0.001	−1.308		
ExtVal	−0.001	0.001	−0.631		
Step 3a (interactions with frequency)				0.272***	0.000
Val × Freq	0.000	0.001	−0.250		
ExtVal × Freq	0.001	0.001	0.728		
Step 4a (form typicality variables)				0.275***	0.003**
TypVal	−0.003	0.001	−2.672*		
TypExtVal	−0.001	0.001	−1.172		
Step 5a (interactions with frequency)				0.277***	0.002
TypVal × Freq	0.002	0.002	1.079		
TypExtVal × Freq	0.002	0.001	1.619		
Step 2b (form typicality variables)				0.276***	0.004**
TypVal	−0.003	0.001	−2.782**		
TypExtVal	−0.001	0.001	−1.227		
Step 3b (interactions with frequency)				0.278***	0.002
TypVal × Freq	0.002	0.002	1.007		
TypExtVal × Freq	0.002	0.001	1.642		
Step 4b (valence variables)				0.277***	−0.001
Val	0.000	0.001	−0.747		
ExtVal	0.000	0.001	−0.479		
Step 5b (interactions with frequency)				0.277***	0.000
Val × Freq	−0.001	0.001	−0.786		
ExtVal × Freq	0.000	0.001	0.307		

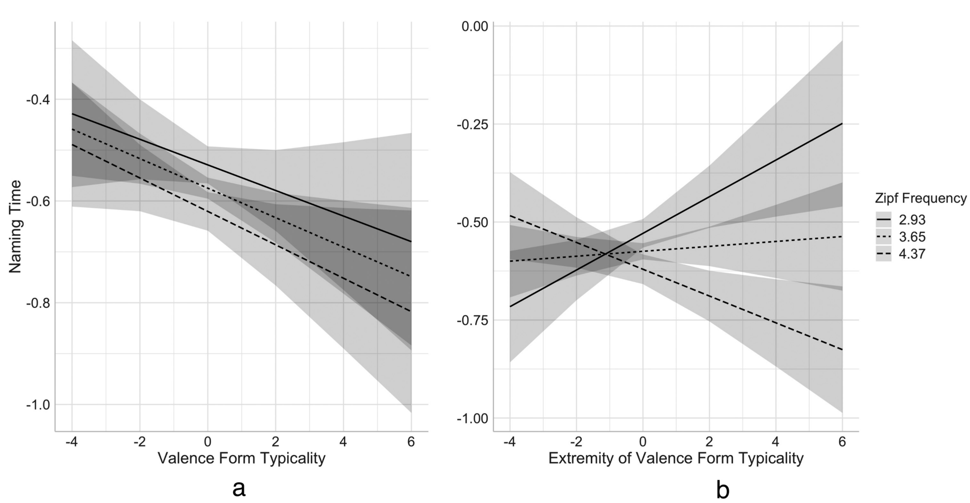
Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency, ELP = English Lexicon Project.
* $p < .05$. ** $p < .01$. *** $p < .001$.

respectively), with more typical forms responded to more quickly. When expressed as a proportion of the mean response latencies in Studies 5 (AELP: 967 ms) and 5b (MALD: 897 ms), these effect sizes translate to a facilitation of approximately 18 and 10 ms of processing time, respectively.

We were unable to replicate reports of valence significantly predicting articulation latencies or accuracy with the ELP data set in Study 6 (cf., Estes & Adelman, 2008a, 2008b; Kuperman et al., 2014). Extremity of valence likewise did not significantly predict performance. Again, this might be due to different sets of words

Figure 10

Added Variable Plot Showing the Relationships Between ELP Naming Latencies and Form Typicality for (a) Valence and (b) Extremity of Valence as a Function of Lexical Frequency After Controlling for the Sublexical and Lexical Predictor Variables



Note. Shaded area shows 95% confidence intervals. ELP = English Lexicon Project.

Table 25
Descriptive Statistics for the Variables in Study 7 (n = 1,211)

Variable	*M*	*SD*
Length	5.30	1.3
OLD	1.90	0.52
PLD	1.70	0.55
Log10 bigram frequency	2.90	0.38
Phonographic neighbors	3.30	4.50
Feedforward consistency	0.90	0.18
Feedback consistency	0.86	0.20
$\text{SUBTLEX}_{\text{US}}$ frequency	3.90	0.74
Age of acquisition	7.90	2.60
Prevalence	2.30	0.25
Concreteness	3.60	1.00
Valence	5.20	1.30
Extremity of valence	1.00	0.86
Form typicality (valence)	−0.13	0.91
Form typicality (extremity of valence)	−0.24	0.91
Hits minus false alarms	0.50	0.13

Note. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance.

and control variables across studies, including arousal and prevalence (Brysbaert et al., 2019). The earlier studies using the ELP data set reported only small proportions of unique variance explained by valence in articulation latencies (e.g., 0.58%, Estes & Adelman, 2008a; 0.2%, Kuperman et al., 2014). Here, form typicality for valence significantly predicted both latencies and accuracy, with more positive forms being produced more quickly yet less accurately, also explaining small proportions of variance (0.2% and 0.4%, respectively). Typicality for extremity of valence did not significantly predict either latencies or accuracy, although it did interact with frequency to predict another 0.3% of variance in latencies.

Study 7 revealed the strongest findings for both emotion and form typicality with respect to recognition memory performance. We were able to replicate reports that more emotional words are better remembered regardless of polarity (i.e., extremity of valence; e.g., Adelman & Estes, 2013; Cortese & Khanna, 2022; see Phelps & Sharot, 2008), explaining 3.3% of variance. Consistent with many previous reports, valence did not significantly predict whether words were recognized or not (see MacMillan et al., 2022; cf., Bowen et al., 2018). Form typicality for valence significantly predicted memory performance, with more typical forms being better remembered, explaining an additional 1.3% of variance, while typicality for extremity of valence did not contribute significantly. However, the latter results might be a conservative estimate as the smaller set of words employed in Study 7 involved a higher proportion of less typical forms for both valence and extremity of valence compared to the other studies (see Table 25).

The present set of findings confirms the overall impression from the empirical literature that emotional valence and extremity of valence are weak and inconsistent predictors of single-word comprehension and production performance, questioning the prominence afforded to these variables in theoretical accounts of lexical processing. Certainly, they are not consistent with the view that affective content automatically and necessarily influences lexical processing due to the grounding of words' meanings in emotional experience (cf. Vinson et al., 2014; e.g., Winter, 2023). This was also the case

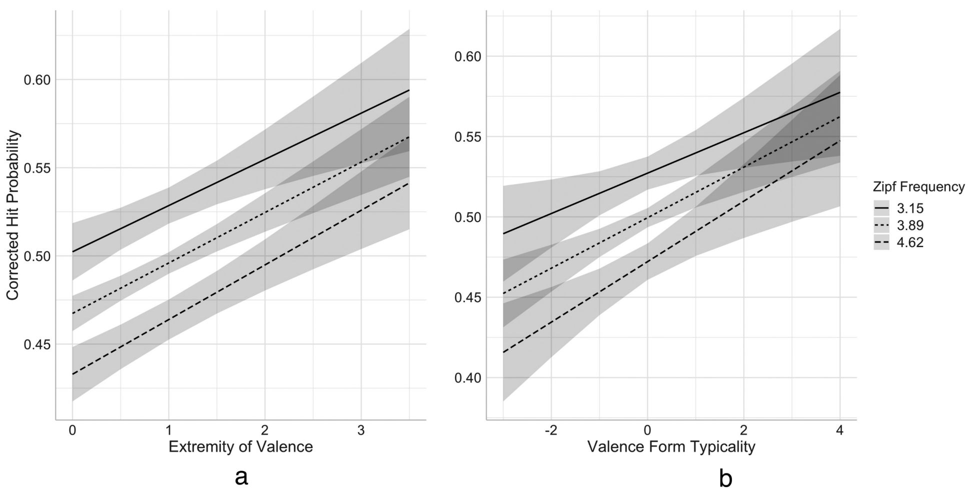
Table 26
Regression Coefficients From Item-Level Analyses of Recognition Memory (n = 1,211)

Model comparison	Estimate	*SE*	*t*	Adjusted R^2	ΔR^2
Step 1 (control variables)				0.278***	
Step 2a (valence variables)				0.311***	0.033***
Val	0.002	0.002	0.952		
ExtVal	0.030	0.004	7.803***		
Step 3a (interactions with frequency)				0.311***	0.000
Val × Freq	0.002	0.003	0.539		
ExtVal × Freq	0.003	0.005	0.604		
Step 4a (form typicality variables)				0.320***	0.009***
TypVal	0.014	0.003	4.184***		
TypExtVal	0.001	0.004	0.250		
Step 5a (interactions with frequency)				0.320***	0.000
TypVal × Freq	0.003	0.005	0.562		
TypExtVal × Freq	−0.004	0.005	−0.763		
Step 2b (form typicality variables)				0.291***	0.013***
TypVal	0.016	0.004	4.431***		
TypExtVal	0.003	0.004	0.931		
Step 3b (interactions with frequency)				0.290***	−0.001
TypVal × Freq	0.004	0.005	0.924		
TypExtVal × Freq	−0.001	0.005	−0.272		
Step 4b (valence variables)				0.320***	0.030***
Val	0.001	0.002	0.218		
ExtVal	0.028	0.004	6.713***		
Step 5b (interactions with frequency)				0.320***	0.000
Val × Freq	0.002	0.003	0.600		
ExtVal × Freq	0.003	0.005	0.647		

Note. Val = valence; ExtVal = extremity of valence; TypVal = form typicality for valence; TypExtVal = form typicality for extremity of valence; Freq = lexical frequency.
*** $p < .001$.

Figure 11

Added Variable Plot Showing the Relationships Between Recognition Memory and (a) Extremity of Valence and (b) Form Typicality for Valence as a Function of Lexical Frequency After Controlling for the Sublexical and Lexical Predictor Variables



Note. Shaded area shows 95% confidence intervals.

for the nonarbitrary statistical relationships we identified between words' emotional content and their surface form features. Interactions with frequency were also weak and inconsistent for both types of predictor variable, and only observable in the written word recognition tasks. As might be expected, form typicality effects were relatively stronger for tasks that required parsing/assembly of phonological/phonetic information (i.e., auditory lexical decision and reading aloud) compared to written lexical decisions that predominantly rely on orthographic activation (Mandera et al., 2020). This mechanism might also explain the result for form typicality in the recognition memory megastudy data, as the procedure involved self-paced, silent reading (i.e., covert articulation) of written words in study lists followed by old/new decisions at test (Cortese & Khanna, 2022).

Implications for Models of Language Processing

While various mechanisms have been proposed to explain how emotional content might affect language and memory processing (e.g., Bowen et al., 2018; Lang et al., 1990; Phelps & Sharot, 2008; Pratto & John, 1991; Unkelbach et al., 2008), there are relatively few such proposals concerning nonarbitrary mappings between form and emotional content. The latter accounts primarily concern auditory processing and assume that negatively valenced words are prioritized due to their alarm signaling properties (e.g., Adelman et al., 2018; Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022). However, this proposal is only partially supported by our findings from Study 5a with the AELP data set. Across both AELP and MALD megastudy data sets, extremity of valence was the stronger and more consistent predictor of auditory lexical decision latencies.

Systematic form–meaning associations tend to occur more frequently among words that are acquired early during language development (Cassani & Limacher, 2022; Monaghan et al., 2014), supporting an important role for statistical learning in bootstrapping language acquisition (e.g., Arciuli, 2017, 2018; Monaghan et al., 2014). In our view, accounts of language and memory processing that incorporate Bayesian statistical inference as a prediction mechanism are well placed to incorporate a priori knowledge about nonarbitrary form–meaning mappings represented in the skilled adult lexicon. Bayesian models have been implemented for speech perception (e.g., Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008), visual word recognition and reading aloud (e.g., Kinoshita, 2015; Norris, 2013), and recognition memory (S. Kim et al., 2017).

The finding of nonarbitrary form–emotion relationships in English words also has implications for the interpretation of brain activation patterns associated with valence and extremity of valence in neuroimaging studies employing single-word comprehension and production tasks (e.g., Kuchinke et al., 2005; Posner et al., 2009; Vigliocco et al., 2014). These studies have frequently interpreted valence-related activation in regions such as the inferior frontal cortex, insula and anterior cingulate as representing grounding in emotional experience or emotional processing. Yet, all three regions have also been consistently implicated in processing of surface form features in neuroimaging studies of perception and production (e.g., Aleman et al., 2005; Blumstein et al., 2005; Minicucci et al., 2013; see Kearney & Guenther, 2019). Hence, it is possible that the activation observed during manipulations of word valence in these studies is due to processing of nonarbitrary form–valence associations rather than "grounding" in affective experience. Future studies should consider testing this alternative hypothesis.

Constraints on Generality

The present study investigated the evidence for sensory form typicality in the English Lexicon. To determine whether sensory form typicality influenced behavioral performance, we employed data sets from megastudies that tested samples of native English-speaking undergraduate students enrolled at universities in the United States (Balota et al., 2007; Cortese & Khanna, 2022), Canada (Tucker et al., 2019), and Singapore (Goh et al., 2020), as well as a crowd-sourced international sample of native English speakers (Mandera et al., 2020). While we would expect these results to generalize to similar tasks conducted in English with native speakers, it remains to be demonstrated whether form typicality for affective content is a statistical property of the lexicons of other languages and, if so, whether it can influence performance on similar tasks in those languages. We have no reason to believe that the results we reported from the behavioral megastudy data sets depend on other characteristics of the participants, materials, or context.

Conclusions

Our findings demonstrate that there are systematic mappings between English words' emotional content and their surface form features. Using behavioral megastudy data, we showed these nonarbitrary relationships significantly influence how words are processed and remembered. However, these effects tended to be as weak and inconsistent across tasks as those for ratings of emotional content, with the strongest effects observed for auditory lexical decision and recognition memory. This might be because emotional information can be more flexibly conveyed through affective prosody independent of word meaning (e.g., Frick, 1985; Kamiloğlu et al., 2020; S. K. Kim & Sumner, 2017). Future studies might also consider exploring nonarbitrary statistical relationships between word forms and ratings of more specific emotional content categories, such as fear, disgust, anger, sadness, and joy (e.g., S. M. Mohammad & Turney, 2013; see Gao, Shinkareva, & Desai, 2022; Gao, Shinkareva, & Peelen, 2022).

References

Adelman, J. S., & Brown, G. D. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review*, *14*(3), 455–459. https://doi.org/10.3758/BF03194088

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535. https://doi.org/10.1016/j.cognition.2013.08.014

Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, *175*, 122–130. https://doi.org/10.1016/j.cognition.2018.02.007

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*(2), 221–228. https://doi.org/10.1093/cercor/bhh124

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, *133*(3), 323–338. https://doi.org/10.1037/0096-3445.133.3.323

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), Article 20160058. https://doi.org/10.1098/rstb.2016.0058

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 634–643. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0135

Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, *59*(5), 920–948. https://doi.org/10.1080/02724980443000782

Arciuli, J., & Cupples, L. (2007). Would you rather 'embert a cudsert' or 'cudsert an embert'? How spelling patterns at the beginning of English disyllables can cue grammatical category. In A. C. Schalley & D. Khlentzos (Eds.), *Mental states, Vol. 2. Language and cognitive structure* (pp. 213–238). John Benjamins.

Arciuli, J., & Monaghan, P. (2009). Probabilistic cues to grammatical category in English orthography and their influence during reading. *Scientific Studies of Reading*, *13*(1), 73–93. https://doi.org/10.1080/10888430802633508

Bakhtiari, G., Körner, A., & Topolinski, S. (2016). The role of fluency in preferences for inward over outward words. *Acta Psychologica*, *171*, 110–117. https://doi.org/10.1016/j.actpsy.2016.10.006

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

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

Béligon, S. (2020). Feeling, emotion and the company they keep: What adjectives reveal about the substantives feeling and emotion. *Lexis*, *15*(1), Article 4322. https://doi.org/10.4000/lexis.4322

Blair, G., Cooper, J., Coppock, A., Humphreys, M., & Sonnet, L. (2022). *estimatr: Fast estimators for design-based inference*. University of California at Los Angeles. https://cran.r-project.org/package=estimatr

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353–1366. https://doi.org/10.1162/0898929054985473

Bowen, H. J., Kark, S. M., & Kensinger, E. A. (2018). NEVER forget: Negative emotional valence enhances recapitulation. *Psychonomic Bulletin & Review*, *25*(3), 870–891. https://doi.org/10.3758/s13423-017-1313-9

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Technical Report C-1). Center for Research in Psychophysiology, University of Florida.

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the $\text{SUBTLEX}_{\text{US}}$ word frequencies. *Behavior Research Methods*, *44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cassani, G., & Limacher, N. (2022). Not just form, not just meaning: Words with consistent form-meaning mappings are learned earlier. *Quarterly Journal of Experimental Psychology*, *75*(8), 1464–1482. https://doi.org/10.1177/17470218211053472

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, *30*(3), 348–369. https://doi.org/10.1016/0749-596X(91)90041-H

Chee, Q. W., Chow, K. J., Yap, M. J., & Goh, W. D. (2020). Consistency norms for 37,677 English words. *Behavior Research Methods*, *52*(6), 2535–2555. https://doi.org/10.3758/s13428-020-01391-7

Connell, L., Lynott, D., & Banks, B. (2018). Interoception: The forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752), Article 20170143. https://doi.org/10.1098/rstb.2017.0143

Cortese, M. J., & Khanna, M. M. (2022). Relating emotional variables to recognition memory performance: A large-scale re-analysis of megastudy data. *Memory*, *30*(7), 915–922. https://doi.org/10.1080/09658211.2022.2055080

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2,897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

de Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

de Zubicaray, G., Arciuli, J., Guenther, F., McMahon, K., & Kearney, E. (2023). Non-arbitrary mappings between size and sound of English words: Form typicality effects during lexical access and memory. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218231184940

de Zubicaray, G., Arciuli, J., Kearney, E., Guenther, F., & McMahon, K. (2023). On the roles of form systematicity and sensorimotor effects in language processing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *49*(3), 431–444. https://doi.org/10.1037/xlm0001201

de Zubicaray, G. I., Arciuli, J., & McMahon, K. (2021). A sound explanation for motor cortex engagement during action word comprehension. *Journal of Cognitive Neuroscience*, *33*(1), 129–145. https://doi.org/10.1162/jocn_a_01640

de Zubicaray, G. I., Kearney, E., Arciuli, J., Guenther, F., & McMahon, K. L. (2023, June 9). *Revisiting form typicality of English sensory words*. https://osf.io/fe4py

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Engelen, J. A. A. (2022). The in-out effect in the perception and production of real words. *Cognitive Science*, *46*(9), Article e13193. https://doi.org/10.1111/cogs.13193

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

Estes, Z., & Adelman, J. S. (2008b). Automatic vigilance for negative words is categorical and general. *Emotion*, *8*(4), 453–457. https://doi.org/10.1037/a0012887

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412–429. https://doi.org/10.1037/0033-2909.97.3.412

Gao, C., Shinkareva, S. V., & Desai, R. H. (2022). SCOPE: The South Carolina psycholinguistic metabase. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01934-0

Gao, C., Shinkareva, S. V., & Peelen, M. V. (2022). Affective valence of words differentially affects visual and auditory word recognition. *Journal of Experimental Psychology. General*, *151*(9), 2144–2159. https://doi.org/10.1037/xge0001176

Goh, W. D., Yap, M. J., & Chee, Q. W. (2020). The Auditory English Lexicon Project: A multi-talker, multi-region psycholinguistic database of 10,170 spoken words and nonwords. *Behavior Research Methods*, *52*(5), 2202–2231. https://doi.org/10.3758/s13428-020-01352-0

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology*, *7*, Article 976. https://doi.org/10.3389/fpsyg.2016.00976

Heise, D. R. (1966). Sound-meaning correlations among 1,000 English words. *Language and Speech*, *9*(1), 14–27. https://doi.org/10.1177/002383096600900102

Huttar, G. L. (1968). Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research*, *11*(3), 481–487. https://doi.org/10.1044/jshr.1103.481

Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods*, *45*(1), 160–168. https://doi.org/10.3758/s13428-012-0242-9

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin & Review*, *27*(2), 237–265. https://doi.org/10.3758/s13423-019-01701-x

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests* (R package Version 0.7.0) [Computer software]. https://CRAN.R-project.org/package=rstatix

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience*, *34*(9), 1214–1229. https://doi.org/10.1080/23273798.2019.1589541

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349–364. https://doi.org/10.1037/0033-295X.99.2.349

Keuleers, E., & Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology*, *68*(8), 1457–1468. https://doi.org/10.1080/17470218.2015.1051065

Kim, S., Potter, K., Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2017). A Bayesian race model for recognition memory. *Journal of the American Statistical Association*, *112*(517), 77–91. https://doi.org/10.1080/01621459.2016.1194844

Kim, S. K., & Sumner, M. (2017). Beyond lexical meaning: The effect of emotional prosody on spoken word recognition. *The Journal of the Acoustical Society of America*, *142*(1), EL49–EL55. https://doi.org/10.1121/1.4991328

Kinoshita, S. (2015). Visual word recognition in the Bayesian Reader Framework. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 63–75). Oxford University Press.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

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

Kuchinke, L., Jacobs, A. M., Grubich, C., Võ, M. L., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: An fMRI study. *NeuroImage*, *28*(4), 1022–1032. https://doi.org/10.1016/j.neuroimage.2005.06.050

Kuhn, M. (2022). *Classification and regression training*. https://CRAN.R-project.org/package=caret

Kuperman, V. (2015). Virtual experiments in megastudies: A case study of language and emotion. *Quarterly Journal of Experimental Psychology (2006)*, *68*(8), 1693–1710. https://doi.org/10.1080/17470218.2014.989865

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

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940. https://doi.org/10.1037/a0030811

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*(3), 377–395. https://doi.org/10.1037/0033-295X.97.3.377

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed:

Importance of word arousal. *Emotion*, *8*(4), 445–452. https://doi.org/10.1037/1528-3542.8.4.445

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology*, *71*(10), 2207–2222. https://doi.org/10.1177/1747021817739834

Liben-Nowell, D., Strand, J., Sharp, A., Wexler, T., & Woods, K. (2019). The danger of testing by selecting controlled subsets, with applications to spoken-word recognition. *Journal of Cognition*, *2*(1), Article 2. https://doi.org/10.5334/joc.51

Louivere, J. L., Flynn, T. N., & Marley, A. A. (2015). *Best-worse scaling: Theory, methods and applications*. Cambridge University Press.

Louwerse, M., & Qu, Z. (2017). Estimating valence from the sound of a word: Computational, experimental, and cross-linguistic evidence. *Psychonomic Bulletin & Review*, *24*(3), 849–855. https://doi.org/10.3758/s13423-016-1142-2

Lumley, T. (2022). *Regression subset selection*. https://CRAN.R-project.org/package=leaps

MacMillan, M. B., Field, H. R., Neath, I., & Surprenant, A. M. (2022). Valence does not affect recognition. *Canadian Journal of Experimental Psychology = Revue canadienne de psychologie experimentale*, *76*(2), 111–121. https://doi.org/10.1037/cep0000275

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior research methods*, *52*(2), 741–760. https://doi.org/10.3758/s13428-019-01272-8

Maschmann, I. T., Körner, A., Boecker, L., & Topolinski, S. (2020). Front in the mouth, front in the word: The driving mechanisms of the in-out effect. *Journal of Personality and Social Psychology*, *119*(4), 792–807. https://doi.org/10.1037/pspa0000196

Minicucci, D., Guediche, S., & Blumstein, S. E. (2013). An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. *Neuropsychologia*, *51*(10), 1980–1988. https://doi.org/10.1016/j.neuropsychologia.2013.06.016

Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In I. Gurevych & Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Vol. 1. *Long Papers* (pp. 174–184). Association for Computational Linguistics.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

Monaghan, P., Chater, N., & Christiansen, M. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition*, *96*(2), 143–182. https://doi.org/10.1016/j.cognition.2004.09.001

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of phonological typicality: Robust coherence and psychological validity. *Mental Lexicon*, *5*(3), 281–299. https://doi.org/10.1075/ml.5.3.02mon

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is English? *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, *369*(1651), Article 20130299. https://doi.org/10.1098/rstb.2013.0299

Murphy, S. C., & Bastian, B. (2020). Emotionally extreme life experiences are more meaningful. *Journal of Positive Psychology*, *15*(4), 531–542. https://doi.org/10.1080/17439760.2019.1639795

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52. https://doi.org/10.3758/BF03193811

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524. https://doi.org/10.1016/j.tics.2013.08.003

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Patil, I. (2021). Visualizations with statistical details: The ‘ggstatsplot’ approach. *Journal of Open Source Software*, *6*(61), Article 3167. https://doi.org/10.21105/joss.03167

Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science*, *17*(2), 147–152. https://doi.org/10.1111/j.1467-8721.2008.00565.x

Pollock, L. (2018). Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods*, *50*(3), 1198–1216. https://doi.org/10.3758/s13428-017-0938-y

Posner, J., Russell, J. A., Gerber, A., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., & Peterson, B. S. (2009). The neurophysiological bases of emotion: An fMRI study of the affective circumplex using emotion-denoting words. *Human Brain Mapping*, *30*(3), 883–895. https://doi.org/10.1002/hbm.20553

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*(3), 380–391. https://doi.org/10.1037/0022-3514.61.3.380

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Reilly, J., Hung, J., & Westbury, C. (2017). Non-arbitrariness in mapping word form to meaning: Cross-linguistic formal markers of word concreteness. *Cognitive Science*, *41*(4), 1071–1089. https://doi.org/10.1111/cogs.12361

Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science*, *31*(1), 157–168. https://doi.org/10.1080/03640210709336988

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

Sereno, J. A., & Jongman, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research*, *19*(6), 387–404. https://doi.org/10.1007/BF01068886

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *Quarterly Journal of Experimental Psychology*, *62*(6), 1115–1122. https://doi.org/10.1080/17470210802618900

Sharpe, V., & Marantz, A. (2017). Revisiting form typicality of nouns and verbs. *Mental Lexicon*, *12*(2), 159–180. https://doi.org/10.1075/ml.17004.sha

Siakaluk, P. D., Newcombe, P. I., Duffels, B., Li, E., Sidhu, D. M., Yap, M. J., & Pexman, P. M. (2016). Effects of emotional experience in lexical decision. *Frontiers in Psychology*, *7*, Article 1157. https://doi.org/10.3389/fpsyg.2016.01157

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*(5), 1619–1643. https://doi.org/10.3758/s13423-017-1361-1

Suárez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighbourhood effects without neighbours. *Psychonomic Bulletin & Review*, *18*(3), 605–611. https://doi.org/10.3758/s13423-011-0078-9

Taylor, I. K., & Taylor, M. M. (1962). Phonetic symbolism in four unrelated languages. *Canadian Journal of Psychology*, *16*(4), 344–356. https://doi.org/10.1037/h0083261

Topolinski, S., Maschmann, I. T., Pecher, D., & Winkielman, P. (2014). Oral approach-avoidance: Affective consequences of muscular articulation dynamics. *Journal of Personality and Social Psychology*, *106*(6), 885–896. https://doi.org/10.1037/a0036477

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The Massive Auditory Lexical Decision (MALD) database. *Behavior Research Methods*, *51*(3), 1187–1204. https://doi.org/10.3758/s13428-018-1056-1

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*(1), 36–49. https://doi.org/10.1037/0022-3514.95.1.36

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

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

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

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Westbury, C., & Moroschan, G. (2009). Imageability × phonology interactions during lexical access: Effects of modality, phonological neighborhood, and phonological processing efficiency. *Mental Lexicon*, *4*(1), 115–145. https://doi.org/10.1075/ml.4.1.05wes

Wilcox, R. R. (2016). *Understanding and applying basic statistical methods using R*. John Wiley & Sons.

Winter, B. (2023). Abstract concepts and emotion: Cross-linguistic evidence and arguments against affective embodiment. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1870), Article 20210368. https://doi.org/10.1098/rstb.2021.0368

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

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yu, C. S., McBeath, M. K., & Glenberg, A. M. (2021). The gleam-glum effect: /i:/ versus /λ/ phonemes generically carry emotional valence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1173–1185. https://doi.org/10.1037/xlm0001017

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. https://cran.r-project.org/package=lmtest

Zipf, G. (1949). *Human behaviour and the principle of least effort*. Addison-Wesley.

Received January 31, 2023
Revision received July 11, 2023
Accepted July 18, 2023 ■