Journal of Memory and Language 150 (2026) 104769

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

That sounds exciting! Emotional iconicity facilitates processing and recall

David M. Sidhu [a,*,1], Pragati Parajuli [a,1], Emiko J. Muraki [b]

[a] Department of Psychology, Carleton University, Canada
[b] Department of Psychology, Western University, Canada

A B S T R A C T

Iconicity in spoken language refers to a perceived resemblance between the form of a word and its meaning. Here we examined emotional iconicity: the relationship between the sound of a word and its emotional arousal. We first examined which language sounds are associated with calmness vs. excitement by collecting emotional arousal ratings of 5,457 nonwords (Experiment 1). This revealed a pattern in which high pitch sounds (e.g., voiceless stops) were associated with excitement while low pitch sounds (e.g., sonorants) were associated with calmness. We used these ratings to identify words whose sounds were iconically congruent (e.g., "shock", containing exciting sounds and with an exciting meaning), incongruent (e.g., "alarm", containing calming sounds and with an exciting meaning) or neutral. In two experiments (Experiment 2 and 3) we found that congruent words were responded to faster, and that incongruent words were responded to less accurately, on a semantic decision task ("Is this word calming or exciting?"). Congruent words were also more likely to be recalled in Experiment 2, but not in Experiment 3 after removing taboo items. However, we found that emotionally iconic words were more likely to be recalled in a mega study dataset of 1,551 items. We also examined whether individual differences in the experience of emotion moderated the effects we observed. Together these results suggest that iconicity is an important factor in language processing.

Emotional iconicity facilitates processing and recall

In spoken language, iconicity refers to instances in which the sound of a word is perceived as resembling its meaning (for reviews see Dingemanse et al., 2015, 2020; Perniss et al., 2010; Winter et al., 2023). Examples of iconicity can be seen in onomatopoeic words such as "cuckoo", which mimic the real-world sounds they represent. Iconicity is theoretically important because it represents an instance of non-arbitrariness in the connection between form and meaning in language (see Dingemanse et al., 2015). That is, a way in which form can connect with meaning through resemblance, as opposed to convention alone.

Crucially, iconicity can also exist crossmodally, when the sound of a word is perceived as resembling some non-auditory property. For example, research has shown that individuals perceive certain language sounds (i.e., *phonemes*) as being evocative of smallness (e.g., /i/ as in "see"), and others as evocative of largeness (e.g., /ɑ/ as in "saw"; Knoeferle et al., 2017; Sapir, 1929; Thompson & Estes, 2011). This leads to a word like "teeny" seeming iconic through its phoneme-size association, even though it does not mimic a sound. Crossmodal iconic associations have also been demonstrated for properties such as shape (Köhler, 1929), colour (see Johansson et al., 2020), personality (Sidhu et al., 2019) and even taste (Gallace et al., 2011). The present study focuses on emotional iconicity (see Aryani et al., 2018; Aryani & Jacobs, 2018; Rummer et al., 2014; Schmidtke et al., 2025; Yu et al., 2021), in particular associations between phonemes and different levels of excitement, and the implications of this for language processing.

Emotional iconicity

Emotion is typically categorized along two axes: valence (referring to how pleasant or unpleasant something is) and arousal (referring to how exciting or calming something is; see Warriner et al., 2013). Warriner et al. (2013) collected ratings of these two dimensions for over 13,000 words, and these ratings are the standard measure of emotion in the literature. Rummer et al. (2014; see also Yu et al., 2021) demonstrated that the vowel /i/ (as in "beet") is associated with positive valence, and that the vowel /o/ (as in the German "brot") is associated with negative valence. Their interpretation was that the articulation of /i/ involves contraction of the same muscle that is used to smile (i.e., zygomaticus major), while /o/ involves the antagonistic muscle (i.e., orbicularis orbis muscle). This was further supported by Körner and Rummer (2022) who found that /i/ was rated more positively than /o/ or /y/ (a front rounded vowel; as in the French "lune"). Note that /y/ and /i/ share a higher pitch, while only /i/ involves activation of the muscle used to smile. This provided elegant support for the role of articulation (vs. pitch) in

* Corresponding author at: 1125 Colonel By Drive Ottawa, Ontario K1S 5B6, Canada.
E-mail address: david.sidhu@carleton.ca (D.M. Sidhu).
[1] These authors are sharing first authorship.

https://doi.org/10.1016/j.jml.2026.104769
Received 31 July 2025; Received in revised form 1 May 2026; Accepted 14 May 2026
Available online 25 May 2026

valence sound symbolism. Notably, these associations are observable in language. Yu et al. (2021) found that in pairs of monosyllabic English words that contained either /i/ or /ʌ/ (e.g., "gleam" vs. "glum"), the word with an /i/ tended to have a more positive meaning.

Our particular interest is in associations between phonemes and arousal. Aryani et al. (2018) examined arousal sound symbolism by having German-speaking participants listen to sets of nonwords and rate them on an arousal scale. They then extracted acoustic features from the nonwords and developed an acoustic model to predict significant variance in perceived arousal. Nonwords with a higher fundamental frequency (i.e., a higher pitch) were judged as more arousing. This is consistent with a general tendency to associate higher pitch vocalizations with greater states of arousal in the vocalizer (i.e., the pitch rule; see Schwartz et al., 2024). Aryani et al. (2018) also found a general pattern in which words with a lower intensity (less acoustic energy) were rated as more exciting. To follow up on this, they mapped the presence of certain phonemes to sound intensity and correlated those phonemes with perceived arousal. They found that short vowels (e.g., /ɪ/ as in "bit"), stops (i.e., consonants involving a complete stoppage in airflow), and hissing sibilants (e.g., /s/ as in "piss") were all predictive of a word being more emotionally arousing. Conversely, long vowels (e.g., /u/ as in "who") and voiced constants (e.g., /g/ as in "gas") were predictive of a word being rated as less emotionally arousing.

Notably, several of the patterns discovered by Aryani et al. (2018) are consistent with the processing of emotional communication in general (e.g., the aforementioned pitch rule). In addition, Aryani et al. (2018) suggest that short vowels might be judged as more arousing because they resemble the rapid breathing that comes along with states of high arousal. In addition, they note that the sounds made by animals in aggressive situations tend to be shorter in duration. Thus, sound properties (linguistic and non-linguistic) may be automatically interpreted and processed as cues to emotion, allowing words to have an iconic relationship with emotional meanings.

A variety of neuroimaging studies have shown that activity in regions of the brain responsible for processing emotion are affected by acoustic properties (see Frühholz et al., 2016). An fMRI study by Aryani et al. (2019) had participants listen to words that sounded either high arousing or low arousing while controlling for lexical (i.e., semantic) arousal. Their findings showed that words that sounded high arousing compared to low arousing shared substantial processing networks with brain regions that process other types of (non-word) affective sounds. Their findings provide neuroimaging evidence of a unifying neural system for affective sound processing, suggesting there are evolutionary mechanisms linking sound and emotion. Additionally, there is also some evidence that these associations extend beyond humans, and can be seen in perceptions of animal calls (Schwartz et al., 2024). These results further suggest an evolutionary basis for these emotional associations to sound.

Effects of iconicity

Beyond demonstrating the existence of iconicity, researchers have begun to explore its effects on cognition. Numerous studies have shown that the first learned words tend to be high in iconicity (Perry et al., 2015; Sidhu et al., 2022; Winter, Perlman, et al., 2022). Researchers have suggested that the resemblance between sound and meaning might make it easier for infants to connect the two when learning language (Imai & Kita, 2014; for a review see Nielsen & Dingemanse, 2021). This has been supported by laboratory studies showing that infants are better able to generalize a new word to novel instances if it is iconic (Imai et al., 2008).

Other work has explored the possibility that iconicity might facilitate word processing. Meteyard et al. (2015) compared the processing of iconic words (primarily onomatopoeia) to non-iconic words. They found that iconic words were responded to more accurately on reading aloud and auditory lexical decision ("Is this a real word?") tasks. Their interpretation was that iconic words enjoy stronger connections between semantic and phonological information, facilitating retrieval. Sidhu et al. (2020) followed this up with a larger set of items and found that iconic words were processed faster and more accurately on a visual lexical decision task, and faster on a phonological lexical decision task (i.e., "Is the sound of this a real word?"; thus pseudohomophones like "brane" also elicit a "yes" response). Importantly, Sidhu et al. included onomatopoeia, crossmodally iconic words (judged by the authors to "not directly imitate their meanings", p. 168), and non-iconic words. The original analyses did not distinguish between these two types of iconicity. However, subsequent analyses (see Online Supplementary Material: https://osf.io/ptfqy/files/8du9j) revealed that crossmodally iconic words were responded to marginally faster in Experiment 1 and significantly faster in Experiment 2. Though, in both cases this effect was smaller than what was observed for onomatopoeia.

Sidhu and Pexman (2022) explored the processing of crossmodally iconic words by operationalizing a specific instance, namely *size iconicity*. In this study, words were considered congruent if the size associations of their phonemes matched their true sizes, and incongruent if they mismatched. For instance, words like "flea" (i.e., small vowel/small object) and "boat" (i.e., large vowel/large object) were considered congruent; words like "ship" (i.e., small vowel/large object) and "wasp" (i.e., large vowel/small object) were considered incongruent. Participants took part in a semantic decision task in which they categorized words based on their physical size as quickly as possible. Words were presented visually (Experiment 1) or auditorily (Experiment 2). The researchers found no evidence of crossmodal iconicity facilitating processing.

However, there has been evidence of facilitation for *emotional* iconicity. In particular, Aryani and Jacobs (2018) used a semantic decision task in which German-speaking participants categorized words as "exciting" or "calming". They used the results of Aryani et al. (2018) to compute a "sound score" for candidate words, quantifying the arousal associations of the sounds they contained. Words either contained exciting or calming sounds, making them congruent or incongruent with their exciting or calming meanings. The authors found that congruent words were processed faster and more accurately. This suggests that unlike size iconicity, emotional iconicity may affect word processing. The authors interpreted this effect as arising from emotional cues provided by a word's sound. That is, when a word contains acoustic features that convey emotional cues (e.g., high pitch), this may facilitate (or interfere with) a decision about the emotional content of the word.

There are thus at least two ways in which we can understand the mechanism underlying emotional iconicity. One is that iconic words in general enjoy special links between their sound and meaning (Meteyard et al., 2015; Sidhu et al., 2020). Another is that the emotion associations of a word's sound can facilitate decisions about a word's affective content (Aryani & Jacobs, 2018). Aryani and Jacbos (2018) suggest that when categorizing a congruent word, individuals benefit from two sources of information, the word's meaning and its sound.

Researchers have also begun to explore other downstream consequences of iconicity. In particular, Sidhu et al. (2023) examined effects of iconicity in recognition memory. They found that iconic words tended to have a higher hit rate (i.e., correctly categorized as previously seen) and a higher false alarm rate (i.e., incorrectly categorized as previously seen). In addition, participants set a lower response criterion for iconic words (i.e., were more willing to respond that they had seen a word before if it was iconic). This study did not distinguish between different kinds of iconicity. One goal of the present study is to examine effects of emotional iconicity in particular on memory. Emotionally arousing stimuli tend to be better remembered (e.g., Cortese & Khanna, 2022; Kensinger & Corkin, 2004), making memory an especially interesting domain in which to test emotional iconicity. By studying recall, we explore iconicity in a context in which the effect cannot be explained via a facilitated fast-paced decision (cf., when quickly categorising a word on its emotion).

Finally, it is important to draw a distinction between iconicity and systematicity (Dingemanse et al., 2015). While iconicity refers to a resemblance between form and meaning, systematicity refers to large-scale statistical patterns in a language among similar words. For example, Reilly et al. (2012) demonstrated that there are systematic differences in the forms of abstract and concrete words, with the latter tending to be shorter. Importantly, on a semantic decision task (abstract vs. concrete), participant responses were more accurate for words with typical forms. More recently, de Zubicaray et al. (2023) examined the extent to which a variety of word form features predicted their valence and valence extremity. They found that typicality (in particular, the degree to which forms were typical of words with extreme valence) affected responses on lexical decision and recognition memory tasks (see also de Zubicaray et al., 2024).

Schmidtke and Conrad (2018) approached the systematicity of emotion in language by calculating the sublexical affective value (SAV) of sub-syllabic units in German. Sub-syllabic units were syllable onsets, nuclei or codas. They then calculated the average valence and arousal of all words containing a given unit. Importantly, they showed that individuals are sensitive to the SAV of these units. For instance, Conrad et al. (2022) found a larger N400 (indicative of semantic processing difficulty) when a word's SAV was incongruent with its meaning. Schmidtke and Conrad (2018) also found that participants were faster to detect letters in a word if those letters had a high arousal SAV.

Crucially, iconicity and systematicity are orthogonal properties. That is, systematic patterns may or may not be iconic (see Sidhu, 2025). To illustrate, Winter and Perlman (2021) demonstrated that adjectives for small sizes tended to contain the phoneme /i/, which individuals perceive as resembling smallness (making it an example of *iconic*-systematicity). Conversely, de Zubicaray et al. (2024) found that words with larger referents tend to have (among other features) more linguodental phonemes (e.g., /θ/ as in the onset of "thank"). This is not a feature that has been shown to be perceived as resembling largeness, making it an example of systematicity alone. Complicating matters is the possibility that systematicity may be an example of the lexicon encoding iconic relationships. Indeed, Schmidtke and Conrad (2025) consider SAV as an example of iconicity permeating the lexicon. Given the observation that SAVs can affect processing, an important consideration for any effects of iconicity on processing is the extent to which systematicity plays a role.

Individual differences in emotional iconicity

A secondary goal of the present study was to leverage individual differences to explore the orthogonal question of whether emotional iconicity involves the experience of emotion. That is for instance, whether phonemes evocative of high arousal elicit this emotion experientially. Embodied theories of concept knowledge (i.e., theories in which meaning involves sensorimotor properties) propose that the activation of emotional experience also contributes to richer semantic processing (Barsalou, 1999; Pexman, 2012). When congruent, emotional experience elicited through word sound *and* word semantics may therefore enrich processing. But when incongruent, conflicting sources of this information may result in slower processing in order to reconcile conflicting information or less accurate processing if one source of information is privileged over another. If emotion iconicity has an effect via experienced emotion, then individual differences in the experience of emotion might serve as moderators of emotional iconicity.

Past research has found that individual traits such as anxiety and need for affect (i.e., tendency to seek out emotional experience) are differentially sensitive to emotional sounds and emotional semantics. For example, Simon et al. (2017) found that individuals with social anxiety disorder showed greater amygdala activity in response to loud angry voices (cf. Peschard et al., 2017). In language processing, studies have found that individuals with higher trait anxiety tend to read negative words more quickly compared to neutral words (Knickerbocker et al., 2015) and people with greater need for affect read positive words more slowly compared to neutral and negative words (Lei et al., 2023). In Experiment 2 we examined whether individual differences on a measure of General Anxiety Disorder (Spitzer et al., 2006) would moderate effects of emotional iconicity. In addition, we also examined whether current feelings of anxiety (i.e., state as opposed to trait anxiety) would serve as a moderator (Spielberger, 1983). In Experiment 3, we explored individual differences at the level of personality, by measuring participants on the Emotionality personality factor from the HEXACO model of personality (Lee & Ashton, 2008). Individuals high on Emotionality experience greater fear and anxiety.

We also consider individual differences related to how iconicity arises more generally from phonological representations. It is notable that Aryani and Jacobs (2018) found an effect of iconicity even when stimuli were presented visually. Their interpretation was that visually presented stimuli still activated phonological representations of words, allowing sounds to have an effect even in the absence of actual auditory input. This is consistent with observed effects of phonology in visual word processing (for a review see Cui et al., 2025). It is also notable that many of the studies on emotional sound symbolism have been conducted in German, which Schmidtke and Conrad (2025) note has especially consistent grapheme-sound conversion rules. If the activation of phonological representations is an important part of the mechanism underlying iconicity effects in word processing, then individuals more sensitive to phonological effects may also be more sensitive to all forms of crossmodal iconicity. Inner voice, the tendency to experience thought as language that is "heard", has been associated with increased phonological interference effects in a speeded word-picture verification task (Roebuck & Lupyan, 2020), suggesting that people with a more salient inner voice activate more robust phonological representations of words. Therefore, in Experiment 3 we also included a measure of the tendency to experience an inner voice (i.e., the Verbal scale in the Internal Representations Questionnaire; Roebuck & Lupyan, 2020) to test the extent to which the experience of inner voice moderated any effects of iconicity.

Present study

We begin with the first large-scale characterization of phoneme-arousal associations (i.e., arousal sound symbolism) in English speaking participants, using thousands of nonword ratings to identify the phonemes that are associated with high and low arousal (Experiment 1). We also explore whether these associations exist as systematic patterns in the lexicon. Complementing previous work (i.e., Aryani et al., 2018; Schmidtke & Conrad, 2018), and to facilitate subsequent experiments, these analyses were carried out at the level of phonemes, rather than acoustic properties or sub-syllabic units.

We then use the phoneme associations discovered in Experiment 1 to select stimuli for a semantic decision task in Experiments 2 and 3. Through these experiments we attempted to replicate the effects of Aryani and Jacobs (2018) in an English-speaking population, as well as build on this study in several ways. In addition to words whose sounds are congruent or incongruent with their meanings, we also included words with a neutral sound. This will allow us to determine whether any effects arise due to facilitation of congruent words, and/or interference of incongruent words.

We also examined whether congruent words would be better remembered. Instead of a recognition memory task, in which responses can be affected by factors other than recollection (e.g., response criterion, familiarity; see Uner & Roediger, 2022) we used a recall task. To our knowledge, this was the first time that the effect of iconicity on recall has been measured. In Experiments 2 and 3, we also consider the effects of individual differences related to the experience of emotions and inner voice. However, note that these investigations were exploratory because the sample sizes in Experiments 2 and 3 were based on the primary research questions, and thus our ability to robustly test individual difference effects was underpowered. Finally, we also explore the effects of emotional iconicity on recall in a mega study dataset of over one thousand items (i.e., the Penn Electrophysiology of Encoding and Retrieval Study; available at: http://memory.psych.upenn.edu/Penn_Electrophysiology_of_Encoding_and_Retrieval_Study; organised by Madan, 2021).

Experiment 1

We began with a large-scale measure of phoneme-arousal associations, as previous measures of phoneme-arousal iconicity have been collected in German-speaking participants, and we intend to test English-speakers. In addition, while Aryani et al. (2018) modelled the relationship in terms of overall acoustic properties, here we model the effect at the level of individual phonemes. To that end, participants rated a set of 5,457 nonwords on an emotional arousal scale.

Method

Participants

There were 107 participants ($M_{\text{Age}} = 41.0$, $SD_{\text{Age}} = 12.1$; 58 males, 49 females) recruited through the platform Prolific (https://www.prolific.com/). They were paid the equivalent of £9.53/hour on average. All participants reported normal or corrected-to-normal vision and English fluency. Four participants were removed for failing an attention check (see below). We also checked whether any participants gave the same response to all items or had an average response time faster than would be meaningful (<200 ms). No participants were removed for either of these criteria.

Materials

Our stimuli consisted of nonword recordings taken from the Auditory English Lexicon Project (Goh et al., 2020). In particular, we used nonwords spoken by an American Female speaker. From the roughly 10,000 candidate nonwords we selected those with three or fewer syllables, 10 or fewer letters, and six or fewer phonemes. This was in order to reasonably allow each phoneme to have an effect on judgments. This resulted in a stimulus set of 5,457 stimuli.

Procedure

Participants took part online through the experiment platform Gorilla (https://gorilla.sc/). Each participant rated one of 55 randomly created stimulus lists, each containing 100 nonwords. We limited the number of nonwords that each participant rated to avoid participant fatigue. Participants were presented with instructions modified from Warriner et al. (2013):

> You will see and hear 100 different invented, meaningless words. Even though they are meaningless, you will use a scale to rate the feeling you get from each word.
> The scale ranges from 1 (excitement) to 9 (calm).
> At one extreme of this scale, the word evokes a feeling of being stimulated, excited, frenzied, jittery, wide-awake, or aroused. When a word completely gives off this feeling, you should choose a rating of 1.
> The other end of the scale is when the word evokes a feeling of being relaxed, calm, sluggish, dull, sleepy, or unaroused. When a word completely gives off this feeling, you should choose a rating of 9.
> The numbers also allow you to describe intermediate feelings of excitement/calm by selecting any of the other numbers.
> If a word feels completely neutral, not excited nor at all calm select the middle of the scale (rating 5).

Nonwords were presented one at a time. Each rating trial began with a 500-millisecond blank screen. Then a nonword recording began playing as the word appeared onscreen. The word remained onscreen for 3000 ms (longer than the duration of any audio file). This was then replaced by a screen which asked the participant to “Please rate the word’s association from 1 (excitement) to 9 (calm).” Participants made their response via keyboard press. Following this, participants pressed the space bar to begin the next trial. Nonwords were presented in a random order. We used this rating scale to align with arousal ratings in Warriner et al. (2013), which was to be used in subsequent experiments. However, because it is more intuitive, scores in Warriner et al. and the present study are reversed going forward such that higher scores indicate more exciting ratings, and lower scores indicate more calming ratings. After the 100 trials each participant received an attention check which presented a nonword onscreen but whose audio asked participants to press “0”.

We took an approach used elsewhere (Westbury et al., 2018) in which we included many nonwords, though each was only rated on arousal a small number of times (1,400 rated once, 3,459 rated twice, 499 rated three times, and 99 rated four times). Our interest was in the relationship between specific phonemes and ratings of arousal. With this approach, each phoneme was rated many times, in different nonword contexts. Thus, no single nonword determined the results. In total, there were 10,211 ratings for analysis.

Results

Data and analysis code for all experiments can be found here: https://osf.io/ptfqy/. We first transcribed nonwords phonetically, beginning with existing transcriptions in the AELP database. A trained linguist reviewed these transcriptions and made corrections when necessary (for 728 nonwords). Nonwords were coded based on whether they contained each candidate phoneme (see Table 1). We then used the presence of each phoneme as predictors of nonword arousal rating. We excluded a given phoneme from further analysis if it appeared less than 100 times in the dataset (/ʊ/, /ʒ/, /ð/, /ʊə/ were excluded in this way). We also collapsed /ir/, /ʊr/ and /ər/ into a single predictor for all r-coloured vowels. Finally, we included several control variables for each nonword: number of phonemes, number of syllables, and phonological Levenshtein distance (Yarkoni et al., 2008). Because of the high correlation among these variables (r’s > 0.66), they were reduced to a single component (henceforth *Phonological Component*) using principal components analysis (loadings > 0.74). Nonword ratings were standardized within each participant and analyses were conducted at the level of individual ratings.

The data were analyzed using least absolute shrinkage selection operator (LASSO) regression (see Friedman et al., 2010). This analysis finds regression coefficients that minimize the sum of squared residuals and the sum of the absolute values of all coefficients, by multiplying the summed absolute coefficient values by a value lambda, which is determined via cross-validation and included in the model as error. We used an Adaptive LASSO (using the “glmnet” package in R; Friedman et al., 2010) which tunes the value of lambda for each of the predictors separately (see Zou, 2006). We used a common definition of lambda as the largest value of lambda that results in a model with mean cross-

Table 1
Phonemes used in stimuli transcriptions.

Phoneme category	Phonemes
Monophthong vowels	/i, ɪ, ɛ, æ, ə, u, ʊ, ʌ, ɔ, ɑ/
Diphthong vowels	/eɪ, aɪ, oʊ, ɔɪ, aʊ, ɪə, ʊə/
R coloured vowels	/ər, ir, ʊr/
Voiceless stops	/p, t, k/
Voiced stops	/b, d, g/
Voiceless fricatives and affricates	/f, θ, s, ʃ, h, tʃ/
Voiced fricatives and affricates	/v, ð, z, ʒ, dʒ/
Sonorants	/m, n, ŋ, ɾ, l, w, j/

Note. The phonemes /ʊ/, /ʒ/, /ð/, and /ʊə/ were removed as predictors because they appeared less than 100 times in the stimuli. The r coloured vowels /ər/, /ir/, and /ʊr/ were combined into a single predictor.

validated error within one standard error of the minimum (Friedman et al., 2010). LASSO models are conservative with regards to large coefficient values and guard against overfitting, with many predictor coefficients being shrunk to zero. Our Phonological Component control variable was forced into the model by setting its lambda value to 0. This approach has been used in previous similar analyses of sound symbolism (see Sidhu et al., 2021).

The results of this analysis can be seen in Table 2. Positive values indicate a phoneme's presence predicts a more "exciting" rating; negative values predict a more "calming" rating. The results indicate that the presence of /tʃ/, /k/, /ʌ/, /ʃ/, /p/, /i/, /ɪ/ and /t/ makes a nonword sound more exciting, while the presence of /u/, /ɔ/, /l/, /m/, /n/ and /oʊ/ makes a nonword sound more calming. The phonological component control variable had a positive coefficient, indicating that longer and more phonologically distinct words were also rated as more exciting.

Analysis of Warriner et al. norms

Before continuing to Experiment 2, to help constrain our interpretation of future experiments, we examined the extent to which the presence of any phonemes are predictive of arousal in the English lexicon. That is, whether the English lexicon demonstrates systematicity with regards to arousal. To that end, we ran a version of the above analysis predicting standardized arousal ratings of real words in the Warriner et al. (2013) norms. Our predictors were again the presence of each phoneme in a given word. For the sake of comparison, we again omitted /ʊ/, /ʒ/, /ð/, and /ʊə/; and collapsed /ir/, /ʊr/ and /ər/ into a single predictor for all r-coloured vowels. We also included number of phonemes, number of syllables, and phonological Levenshtein distance (Yarkoni et al., 2008) as control variables. As with the nonwords, there was a high correlation among these variables (r's > 0.79), and so they were reduced to a single Phonological Component (loadings > 0.89). The analysis included a total of 12,438 items.

All phoneme predictors were shrunk to zero in this model, indicating that none of them were a significant predictor of arousal ratings. As was the case with nonwords in Experiment 1, the Phonological Component was a significant positive predictor ($b = 0.11$), indicating that longer and more phonologically distinct words tend to be more exciting. We also ran versions of this analysis on verbs (n = 1,673), adjectives (n = 2,860) and nouns (7,830) alone. All phoneme predictors were shrunk to zero, with the exception of /ŋ/ in the prediction of adjectives ($b = 0.03$).

We also conducted a version which only included words at the extreme ends of arousal (i.e., >1 SD above or below the mean; n = 3,946). With this approach 11 phoneme predictors are retained: /v, ɔɪ, ʃ, h, ɪ, z, dʒ, s, eɪ/ as predictors of high arousal and /oʊ, l/ as predictors of low arousal. See Online Supplementary Material for coefficients. While the results are sensible, we are hesitant to interpret them given the possibility that they are an artifact of eliminating the middle of the dependent variable scale. In addition, only four predictors overlap with those discovered in nonwords.

Table 2
Results of the LASSO analysis predicting nonword rating.

Predictor	Coefficient	Nonwords Containing
/tʃ/	0.240	370
/k/	0.15	2,023
/ʌ/	0.059	646
Phonological component	0.058	
/ʃ/	0.048	499
/p/	0.035	1,800
/ɪ/	0.018	2,146
/i/	0.017	1,324
/t/	0.013	2,538
/ɔ/	−0.028	577
/u/	−0.049	1,409
/oʊ/	−0.057	1,582
/n/	−0.087	2,499
/m/	−0.1	1,928
/l/	−0.105	2,725

Note. Positive coefficients predict an exciting rating; negative coefficients predict a calm rating. Candidate predictors which had their coefficients shrunk to zero were: /æ, ɛ, ɑ, ə, eɪ, ɑɪ, ɔɪ, ɑʊ, ɪə, b, d, dʒ, f, g, h, ɾ, s, v, w, z, θ, ŋ, j/ and the single predictor for the r-coloured vowels.

Discussion

This experiment revealed an association between 14 phonemes and high or low emotional arousal. This is the first demonstration of phoneme-arousal associations (i.e., arousal sound symbolism) using a large-scale set of nonword ratings. Thus, it serves as the first quantification of arousal sound symbolism, done at the phoneme level (see Aryani et al., 2018 for a demonstration at the acoustic level). Importantly, penalized regression is conservative with regards to over-fitting. Indeed, 28 other phoneme predictors had their coefficients "shrunk" to zero. Thus, the resulting coefficients are not merely a product of the large sample size. The general pattern in our results was that voiceless stops (i.e., /p/, /t/ and /k/) were associated with excitement, while nasal sonorants (i.e., /l/, /m/ and /n/) were associated with calmness. Further, the sibilants /ʃ/ and /tʃ/ were both also associated with excitement. In addition, the shorter vowels /ɪ/ and /ʌ/ were associated with excitement, while the longer vowels /oʊ/, /ɔ/ and /u/ were associated with calmness. The one exception to this pattern was that the long vowel /i/ was associated with excitement. This could suggest a different underlying mechanism for this phoneme, perhaps based on pitch. Notably, these results are consistent with those of Aryani et al. (2018) in a sample of German-speakers. Importantly, these associations do not appear to be learned from systematicity, as they were not present in a set of English words rated on their emotional arousal, at least when the entire range of emotional arousal is examined. Next, we used the coefficients we generated from nonword ratings to select real words whose sounds are congruent, neutral or incongruent with their meanings, to be used in a semantic decision and recall task.

Experiment 2

In Experiment 2, we examined whether participants would respond faster and more accurately during semantic processing to words whose sounds were congruent with their meanings in terms of emotional arousal. In addition, we examined whether congruent words would be more likely to be recalled. We also explored whether individual differences in anxiety would moderate any of these effects.

Method

Participants

Participants were 53 undergraduate students from Carleton University (31 female, 18 male, 2 non-binary, one trans male, one unknown; $M_{\text{Age}} = 19.9$, $SD_{\text{Age}} = 3.8$). We based our sample size on Brysbaert and Stevens (2018) suggestion of 1,600 observations per cell. Here we had 40 items per cell (see below), thus leading to the target of 40 participants. After reaching 40 participants we continued collecting data based on convenience sampling in case any participants needed to be excluded. Participants received course credit. All participants reported fluency in English and normal to corrected normal vision.

Materials

To develop a list of stimuli for this study, we began with the 13,915 words from the Warriner et al. (2013) emotional word rating norms. Henceforth we refer to each word's arousal rating in these norms as its *meaning score*. We quantified the emotional arousal of each word's *sound* using the coefficients generated in Experiment 1 (see Table 1). Each word was transcribed into its component phonemes. We summed the coefficients for each phoneme in a given word and then divided this by the number of phonemes in the word. Phonemes without an associated coefficient were treated as having a coefficient of 0. See Fig. 1. Henceforth we refer to this as a word's *sound score*.

We used meaning and sound scores to select our items. Meaning and sound scores were standardized, and scores above 1.25 were considered exciting, while scores below −1.25 were considered calming, and sound scores between −0.25 and 0.25 were considered neutral. These thresholds were a trade-off between being extreme enough to be considered exciting or calming, while also providing sufficient candidates for matching. Using meaning and sound scores, we selected 40 words in each of the following groups, words with an exciting sound with exciting meaning, calming sound with exciting meaning, neutral sound with exciting meaning, exciting sound with calming meaning, calming sound with calming meaning, and neutral sound with calming meaning. Words in each group were matched (p's > 0.11; see Online Supplementary Material for a list of t-tests with a p value < 0.40) on number of letters, number of phonemes, orthographic Levenshtein distance (Yarkoni et al., 2008), concreteness (Brysbaert et al., 2014), subtitle word frequency (Brysbaert & New, 2009), and valence (Warriner et al., 2013). In addition, groups of the same meaning or sound type (e.g., words with "exciting meaning and calming sound", "exciting meaning and exciting sound" are of the same meaning type) were matched on the dimension that they share (e.g., all groups with an exciting meaning were matched on their meaning scores). See Table 3 for example words, and Table 4 for means of control variables in each word type.

We also included two measures of participant anxiety. To measure state anxiety, we used the state items from the State-Trait Anxiety Inventory (Spielberger, 1983). This consists of 20 statements (e.g., "I feel tense") that participants rate in terms of their endorsement on a scale from 1 ("Not at all") to 4 ("Very much so"). This was followed by the General Anxiety Disorder 7 questionnaire which consists of seven items (e.g., "Feeling nervous, anxious or on edge"; Spitzer et al., 2006). Participants self-rated how often they felt the specific way on a scale from 0 ("Not at all") to 3 ("Nearly every day"). This was designed to measure how often the participant exhibited each anxiety symptom over the last two weeks and is used to identify probable cases of generalized anxiety disorder.

Procedure

The study was conducted in person using PsychoPy and Qualtrics experimental platforms. Participants first completed a semantic decision task using PsychoPy. Words were presented on screen one at a time, and participants were instructed to decide whether each word was emotionally exciting or calming. The definitions of these categories presented to participants were modified from the Warriner et al. (2013) norms. Exciting words were described as evoking the feeling of being stimulated, excited etc.; calming words were described as evoking the feeling of being relaxed, calm etc.

Participants made their responses via keyboard press, using the keys 'P' and 'Q'. Key assignment was counterbalanced across participants. Participants had two sets of six item practice trials, one untimed and one timed, before beginning the task proper. Each trial began with a fixation

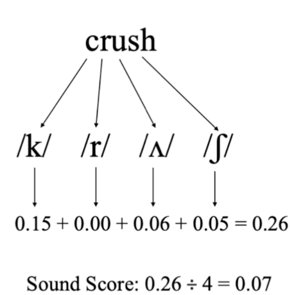

Fig. 1. Example calculation of a word's sound score. Positive values indicate more exciting sounds.

Table 3
Example items of each word type.

Meaning type	Exciting sound	Neutral sound	Calming sound
Exciting Meaning	cocktail, shock	bang, voyage	alarm, ammo
Calming Meaning	cabbage, park	beige, swab	lily, snail

Note. Meaning is based on ratings in Warriner et al. (2013). Sound is based on the average coefficient (from Experiment 1) of each phoneme in a word.

Table 4
Mean (SD) of matching variables for each word type.

Dimension	CMCS	CMES	CMNS	EMCS	EMES	EMNS
Number of letters	5.72 (1.65)	5.65 (1.59)	5.58 (1.77)	5.75 (1.68)	5.80 (1.64)	5.95 (1.58)
Number of phonemes	4.60 (1.34)	4.42 (1.34)	4.72 (1.55)	4.50 (1.41)	4.65 (1.33)	4.80 (1.47)
OLD	2.03 (0.72)	2.03 (0.71)	2.01 (0.66)	2.21 (0.80)	2.11 (0.79)	2.12 (0.68)
Concreteness	3.78 (1.04)	3.95 (0.88)	3.71 (0.98)	3.65 (0.98)	3.63 (0.90)	3.65 (1.01)
Frequency	2.28 (0.70)	2.24 (0.68)	2.27 (0.72)	2.34 (0.63)	2.43 (0.71)	2.36 (0.75)
Valence	5.29 (0.87)	5.47 (0.85)	5.35 (0.85)	5.28 (1.29)	5.35 (1.26)	5.40 (1.25)
Arousal	2.83 (0.19)	2.85 (0.16)	2.84 (0.19)	5.71 (0.36)	5.77 (0.44)	5.78 (0.30)
Sound score	−0.04	0.05	−0.00	−0.04	0.05	−0.00

Note. CMCS, for example, is Calming Meaning and Calming Sound.

cross for 1.5 seconds. This was then replaced by a word which remained onscreen for 1.5 s or until a response was made. This was the same time limit used by Aryani and Jacobs (2018). Participants were instructed to respond as quickly and as accurately as possible. A response from the participant began the next trial. The presentation of words was randomized.

Following the semantic decision task, participants were given a filler task in which they solved 10 addition problems. Following this, they took part in a free recall task in which they were given five minutes to recall as many words as they could from the semantic decision task. They typed their answers into a response window.

Participants were then asked to fill the Spielberg State-Trait Anxiety Index (Sydeman, 2018) and the General Anxiety Disorder 7 questionnaire (Spitzer et al., 2006). The participants were made aware that the questions were only being used for research purposes, and not for diagnosis. Finally, they were also asked to complete a demographic questionnaire including age, gender, languages spoken and English fluency. The recall task and questionnaires were completed through a desktop version of Qualtrics.

Results

All analyses were performed using the R statistical software (R Core Team, 2021), using the packages "lme4" (Bates et al., 2015) and "afex" (Singmann et al., 2016). Analyses consisted of mixed effects models. In the analyses, responses were considered correct in the semantic decision task if the response selection (e.g., exciting) matched the meaning type as defined by the Warriner et al. (2013) arousal norms, although we acknowledge that this operationalization of accuracy does not consider the subjectivity of affective judgements, which may vary across individuals (see Aryani & Jacobs, 2018).

Reaction time

In the analysis of reaction time we removed all incorrect responses (29.83% of trials), trials on which participants did not respond before the deadline (4.24% of trials) and trials with response times less than 200 ms (0.04% of trials). Our analysis consisted of a linear mixed effects model. The dependent variable was response time in milliseconds. Our predictor of interest was word type, based on the congruence between a word's meaning and its sound. The levels of word type were congruent (exciting meaning/exciting sound, calming meaning/calming sound), incongruent (exciting meaning/calming sound, calming meaning/exciting sound) and neutral (exciting meaning/neutral sound, calming meaning/neutral sound). Word type was dummy coded with congruent words as the reference level. We also included subtitle word frequency (Brysbaert & New, 2009), number of letters, orthographic Levenshtein distance (Yarkoni et al., 2008), concreteness (Brysbaert et al., 2014), valence, valence extremity (i.e., absolute distance from midpoint) and arousal (Warriner et al., 2013), as control variables. All continuous variables were standardized. In addition, we included random subject and item intercepts, as well as a random subject slope for word type. In order to deal with convergence issues, here and in all analyses, we attempted the following remedies: increasing the number of iterations, switching to the "bobyqa" optimizer, removing the correlation between the random subject intercept and slope, individually removing the random slopes (beginning with the one with the lowest associated variance), and individually removing the random intercepts (beginning with the one with the lowest associated variance; see Brauer & Curtin, 2018; Meteyard & Davies, 2020).

Our results indicated that congruent words were responded to significantly faster than incongruent words (b = 22.11, p = .008). However, this effect was not significant when comparing to neutral words (b = 15.78, p = .053). See Table 5 and Fig. 2. Responses were also faster for more frequent (b = −9.49, p = .03) and shorter words (b = 26.92, p = .001), as well as words with a more exciting meaning (b = −16.32, p < .001). A supplementary analysis using incongruent words as the reference level found no difference between response times to incongruent and neutral words (b = 6.33, p = .45).

Table 5
Results of the mixed effects linear regression for predicting reaction time in Experiment 2.

Fixed effect	*B*	β (95% CI)	*SE*	*t*	*p*
Intercept	860.208		15.809	54.41	< .001***
Word type (incongruent vs congruent)	22.109	0.05 (0.01, 0.08)	8.197	2.697	.008**
Word type (neutral vs congruent)	15.778	0.03 (0.00, 0.07)	8.093	1.949	.053
Frequency	−9.492	−0.04 (−0.08, −0.01)	4.271	−2.222	.027*
Number of letters	26.924	0.12 (0.05, 0.20)	8.106	3.321	.001**
Number of phonemes	−2.886	−0.01 (−0.07, 0.04)	5.99	−0.481	.63
Orthographic levenshtein distance	−3.893	−0.02 (−0.09, 0.05)	7.864	−0.481	.631
Concreteness	0.299	0.00 (−0.03, 0.03)	3.537	0.085	.932
Valence	2.564	0.01 (−0.02, 0.05)	4.039	0.635	.526
Valence extremity	−6.561	−0.03 (−0.07, 0.01)	4.227	−1.552	.122
Arousal	−16.321	−0.07 (−0.11, −0.04)	3.625	−4.502	< .001***
Random effect	***SD***				
Word intercept	40.246				
Participant intercept	107.399				

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

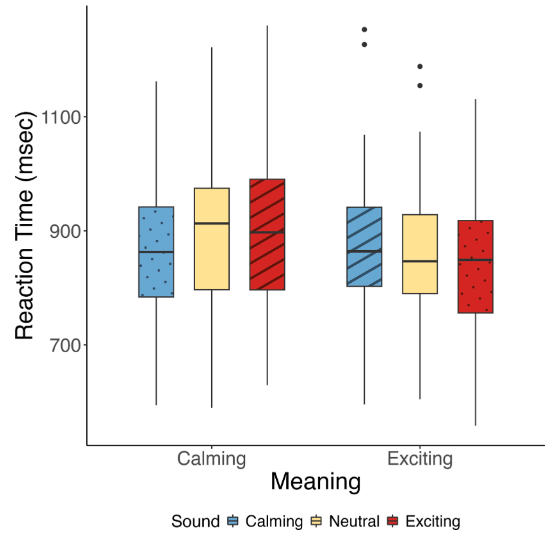

Fig. 2. Average reaction time by meaning and sound type in Experiment 2. *Note.* Dotted bars are congruent pairings, lined bars are incongruent pairings. The plot shows averages by subject.

In a supplementary analysis, we ran a version of the model that also included an interaction between word type and each word's meaning score, to examine whether congruency might affect exciting or calming words differently. This model found no significant interactions (p's > .45). Finally, we ran a version of the model that included an interaction between word type and each participant's state anxiety and generalized anxiety scores. These models also found no significant interactions (p's > . 21).

Accuracy

In the analysis of accuracy we removed trials on which participants failed to respond before the deadline (4.24% of trials)[2] as these may reflect a different process than providing an incorrect response (see Scaltritti et al., 2025). We also removed trials with response times less than 200 ms (0.04% of trials). The analyses followed the same approach as those of reaction time, except that they consisted of a logistic mixed effects regression in which the dependent variable was whether the participant responded correctly on a given trial. Our results indicated that congruent words were responded to significantly more accurately than incongruent words (b = −0.35, p < .001), but not neutral words (b = −0.11, p = .28). See Table 6 and Fig. 3. Responses were also more accurate to more frequent words (b = 0.14, p = .009). A supplementary analysis using incongruent words as the reference level found that they were responded to significantly less accurately than neutral words (b = 0.24, p = .02).

In a supplementary analysis we ran a version of the model that also included an interaction between word type and meaning score, to examine whether the results might only be observed for exciting or calming meanings. This model found no significant interaction (p's > .34). Finally, we ran an exploratory model that included an interaction between word type and each participant's state anxiety and generalized anxiety disorder scores. This model also found no significant interactions (p's > .12).[3]

[2] The pattern for these failures to respond matched the overall pattern in the data, with a failure to respond before the deadline on 2.06% of congruent trials, 2.38% of neutral trials, and 2.30% of incongruent trials.

[3] Interestingly, in a model that did not include an interaction, we found that participants scoring higher in Generalized Anxiety responded more accurately (b = 0.13, p = .01).

Table 6
Results of the mixed effects logistic regression for predicting accuracy in Experiment 2.

Fixed effect	*B*	exp(*B*)	ΔPercentage	*SE*	Wald's *z*	*p*
Intercept	1.031	2.803		0.088	11.645	< .001***
Word type (incongruent vs congruent)	−0.353	0.699	−7.358	0.104	−3.402	< .001***
Word type (neutral vs congruent)	−0.112	0.894	−2.216	0.103	−1.079	.28
Frequency	0.141	1.151	2.92	0.054	2.618	.009**
Number of letters	−0.062	0.940	−1.279	0.103	−0.603	.547
Number of phonemes	0.015	1.015	0.306	0.076	0.195	.845
Orthographic levenshtein distance	0.056	1.058	1.165	0.099	0.569	.569
Concreteness	−0.054	0.948	−1.105	0.044	−1.206	.228
Valence	−0.094	0.948	−1.946	0.051	−1.846	.065
Valence extremity	−0.017	0.983	−0.354	0.053	−0.322	.748
Arousal	−0.061	0.941	−1.265	0.045	−1.348	.178
Random effect	***SD***					
Word intercept	0.563					
Participant intercept	0.361					

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

Recall

These analyses were also conducted using logistic mixed effects regressions, as opposed to ANOVAs which are more common in analyzing recall data.[4] We did this to be able to include lexical semantic control variables. Our dependent variable was whether a given word was recalled correctly by a participant at test. Thus, there were 240 data points per participant, with each being scored as a 0 ("not correctly recalled") or a 1 ("correctly recalled"). One participant was excluded from the analysis due to incomplete data. We took a conservative approach to the data and did not consider a word as correctly recalled if it was misspelled in any way. Our approach to the analysis was identical to that of accuracy. The results indicated that congruent words were significantly more likely to be correctly recalled than incongruent words ($b = -0.41$, $p < .03$), but not neutral words ($b = -0.26$, $p = .15$). See Table 7 and Fig. 4. Participants were also more likely to correctly recall more frequent ($b = 0.24$, $p = .01$), concrete ($b = 0.22$, $p = .005$) and exciting words ($b = 0.34$, $p < .001$). A supplementary analysis using incongruent words as the reference level found that they were no less likely to be recalled correctly than neutral words ($b = 0.15$, $p = .42$).

In a supplementary analysis we ran a version of the model that also included an interaction between word type and meaning score, to examine whether the results might only be observed for exciting or calming meanings. This model found no significant interaction (p's $>$ .31). Finally, we ran an exploratory model that included an interaction between word type and each participant's state anxiety and generalized anxiety disorder scale scores. This model also found no significant interactions (p's $>$ .25).

Discussion

Our results indicated that iconically congruent words were processed faster and more accurately than incongruent words. Congruent words were also more likely to be recalled correctly. These results showed evidence of interference in response accuracy (incongruent words were responded to less accurately than neutral words). None of these effects interacted with individual measures of anxiety. Nevertheless, these results replicate those of Aryani and Jacobs (2018) in an English-speaking population, and extend the behavioural effects of iconicity to recall.

[4] However, the pattern of results are largely identical with a traditional one-way repeated measures ANOVA. It revealed a significant effect of word type (F [1.95, 99.41] $= 11.02$, $p < .001$). Post hoc analyses using estimated marginal means revealed that participants recalled more congruent words than either than incongruent words, $t(51) = 4.49$, $p < .001$, or neutral words, $t(51) = 1.02$, $p = .006$. There was not a significant difference between incongruent words and neutral words, $t(51) = 1.56$, $p = .13$.

It is important to note many words rated highest in excitement in Warriner et al. (2013) are taboo in nature. Though we attempted to exclude as many of these as possible when creating our word lists, several taboo words were included. These were "erect" (congruent), "kinky" (congruent), "nude" (incongruent) and "vibrator" (neutral). Importantly, "erect", "kinky" and "vibrator" were the three words recalled correctly the most often (recalled correctly by 19/53 participants), consistent with previous work showing better memory for taboo items (Jay et al., 2008). Because two of the four taboo words were congruent, we performed an analysis which excluded them, to examine whether this confound was responsible for the observed recall effect. When excluding taboo words from the recall analysis, the significant difference between congruent and incongruent words was not significant ($b = -0.31$, $p = .06$). Removing these words from reaction time and accuracy analyses did not change the results. Nevertheless, in Experiment 3 we ran a replication of the current study after replacing these taboo items.

Experiment 3

We conducted a replication of Experiment 2 after replacing the four taboo items from that experiment. In addition, because we did not find any evidence that measures of anxiety moderated the effect, we included measures of other potential moderators: personality and processing style.

Method

Participants

We used the results of Experiment 2 to conduct an a priori power analysis specifically for our main effect of interest: reaction time on the semantic decision task. This was done using the "mixedpower" package in R to simulate data at different sample sizes based on Experiment 1 (Kumle et al., 2021). This revealed that with 110 participants we would have .99 power to detect the effect of word type for congruent vs. incongruent words, and .88 power for congruent vs. neutral words. We aimed for this sample size but only reached 109 given the time constraints of the semester. These were undergraduate students from Carleton University (68 female, 37 male, three non-binary, one trans man; $M_{Age} = 19.34$, $SD_{Age} = 3.84$). One participant who repeated part of the experiment due to experimenter error was removed. Participants received course credit. All participants reported fluency in English and normal to corrected normal vision.

Materials

Our list of stimuli was the same as in Experiment 2, except that we replaced the four taboo items ("erect", "kinky", "nude" and "vibrator")

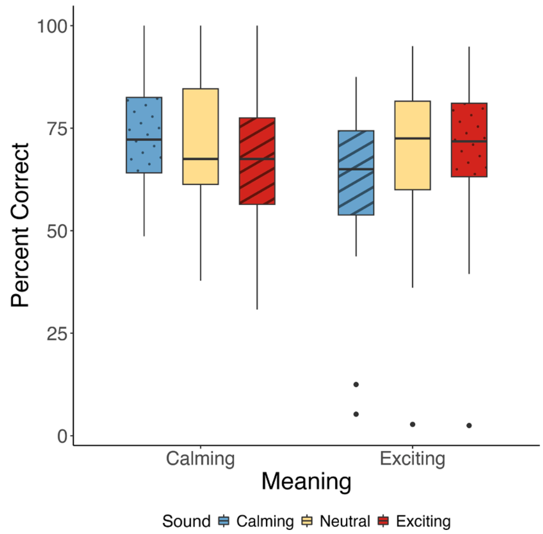

Fig. 3. Average accuracy by meaning and sound type in Experiment 2. *Note.* Dotted bars are congruent pairings, lined bars are incongruent pairings. The plot shows averages by subject.

with words of the same sound and meaning type. We also improved the item set by replacing three words that shared a stem with others on the list ("actuality", "match" and "sale"; which shared stems with "actual", "matching" and "salesman", respectively). After replacement, words in each group were still matched on the same dimensions as in Experiment 2 (p's > 0.12; see Online Supplementary Material for a list of t-tests with a p value < 0.40).

Given the fact that our measures of anxiety did not interact with any relevant predictors, we did not include them. Instead, we included a measure of Emotionality, one factor in the HEXACO model of personality (Lee & Ashton, 2008), which measures the extent to which individuals feel fear and anxiety, and feel empathy and sentimental attachments towards others. In particular, participants rated the extent to which they agree with 10 statements about themselves (e.g., "I feel like crying when I see other people crying."), on a scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). In addition, we included the Internal Representations Questionnaire (*IRQ*; Roebuck & Lupyan, 2020). This consists of 36 statements that together tap into four non-orthogonal "modes of thinking". Our particular interest was in the verbal dimension (i.e., the tendency to experience thought in an inner voice). Participants responded to each statement on a scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree").

Procedure

The procedures for the semantic decision and recall tasks were the same as in Experiment 2. Following these tasks participants completed the IRQ and Emotionality questions.

Results

Our analysis approach was the same as in Experiment 2.

Reaction time

In the analysis of reaction time we removed all incorrect responses (30.10% of trials), trials on which participants did not respond before the deadline (3.81% of trials) and trials with response times less than 200 ms (0.03% of trials). Our results indicated that congruent words were responded to significantly faster than incongruent words (b = 26.17, p < .001), and neutral words (b = 16.25, p = .021). See Table 8 and Fig. 5. Responses were also faster for more frequent (b = −13.99, p < .001) and shorter words (b = 17.60, p = .012), as well as words with a more exciting meaning (b = −18.43, p < .001). A supplementary analysis using incongruent words as the reference level found no difference between response times to incongruent and neutral words (b = −9.92, p = .17).

In a supplementary analysis, we ran a version of the model that also included an interaction between word type and each word's meaning score, to examine whether congruency might affect exciting or calming words differently. This model found no significant interactions (p's > .87). We also ran a version of the model that included an interaction between word type and participants' scores on the verbal subscale of the IRQ. This model found no significant interactions (p's > .19). Finally, we ran an exploratory model that included an interaction between word type and participants' Emotionality. This model found no significant interactions (p's > .09).

Nevertheless, we proceeded with an analysis of the three-way interaction with participant Emotionality, word type, meaning score, and all two-way interactions.[5] This revealed a significant three-way interaction between Emotionality, word type (congruent vs. neutral) and meaning score (b = 8.47, p = .009), see Fig. 6. We followed this up using estimated marginal means and pairwise comparisons of high (SD = 2) and low values (SD = −2) of meaning score and Emotionality. This revealed an effect of emotional iconicity for exciting words for those high in Emotionality (b = −56.80, p = .008), while iconicity for calming words had an effect for those low in Emotionality (b = −43.20, p = .04). In addition, there was an effect of iconic incongruence for exciting words for those low in Emotionality (b = −62.87, p = .005).

We next removed the three-way interaction to allow interpretation of two-way interactions. This revealed a significant two-way interaction between meaning score and Emotionality (b = −3.12, p = .02), see Fig. 7. We followed this up by comparing the simple slope of word meaning for those high (SD = 2) and low (SD = −2) in Emotionality. This revealed a larger effect of word meaning for those high in Emotionality (b = −98.80, p < .001) than those low in Emotionality (b = −48.80, p = .004). No other interactions were significant (p's > .08).

[5] Because our power analysis was only for the main effect of word type, we are likely underpowered to detect three-way interactions. Nevertheless, we conducted exploratory three-way interactions involving word type, meaning score, and all individual difference measures. These will only be reported when the result is significant.

Table 7
Results of the mixed effects linear regression for predicting recall accuracy in Experiment 2.

Fixed Effect	*B*	exp(*B*)	ΔPercentage	*SE*	Wald's *z*	*p*
Intercept	−3.295	0.037		0.146	−22.62	<.001***
Word Type (Incongruent vs Congruent)	−0.412	0.662	−1.311	0.184	−2.236	.025*
Word Type (Neutral vs Congruent)	−0.259	0.772	−0.881	0.181	−1.430	.153
Frequency	0.243	1.275	0.764	0.095	2.55	.01*
Number of Letters	−0.215	0.807	−0.675	0.187	−1.148	.251
Number of Phonemes	0.044	1.045	0.138	0.139	0.316	.752
Orthographic Levenshtein Distance	0.238	1.269	0.748	0.179	1.329	.184
Concreteness	0.225	1.252	0.707	0.079	2.832	.005**
Valence	0.131	1.140	0.412	0.090	1.45	.147
Valence Extremity	−0.015	0.985	−0.048	0.095	−0.162	.872
Arousal	0.336	1.399	1.057	0.080	4.207	<.001***
Random Effect	*SD*					
Word Intercept	0.849					
Participant Intercept	0.458					

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

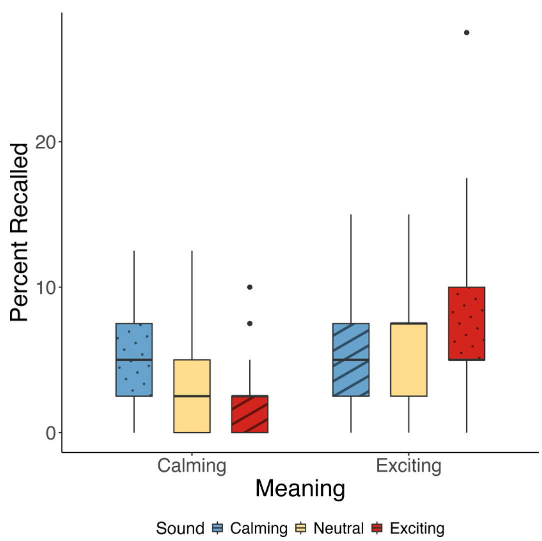

Fig. 4. Average percent recalled by meaning and sound type in Experiment 2. *Note.* Dotted bars are congruent pairings, lined bars are incongruent pairings. The plot shows averages by subject.

Accuracy

In the analysis of accuracy we removed trials on which participants failed to respond before the deadline (3.81% of trials)[6] as well as trials with response times less than 200 ms (3.88% of trials). The analyses followed the same approach as those of reaction time, except that they consisted of a logistic mixed effects regression in which the dependent variable was whether the participant responded correctly on a given trial. Our results indicated that congruent words were responded to significantly more accurately than incongruent words ($b = -0.43$, $p < .001$), but not neutral words ($b = -0.16$, $p = .15$). See Table 9 and Fig. 8. Responses were also more accurate to more frequent words ($b = 0.23$, $p < .001$). A supplementary analysis using incongruent words as the reference level found that they were responded to significantly less accurately than neutral words ($b = 0.27$, $p = .03$).

In a supplementary analysis we ran a version of the model that also included an interaction between word type and meaning score, to examine whether the results might only be observed for exciting or calming meanings. This model found no significant interaction (p's > .35). Finally, we ran separate versions of the model that included an interaction between word type and scores on the verbal subscale of the IRQ or participant's Emotionality. There were no significant interactions (p's > .75).

Recall

Our results indicated that participants were no more likely to recall congruent words than incongruent words ($b = -0.18$, $p = .22$), or neutral words ($b = -0.17$, $p = .25$). See Table 10 and Fig. 9. Participants were more likely to recall a word if it was more frequent ($b = 0.20$, $p = .01$), shorter in letter length ($b = -0.39$, $p = .01$), more orthographically distinct ($b = 0.42$, $p = .004$), more concrete ($b = 0.14$, $p = .03$) and with

[6] The pattern for these failures to respond match the overall pattern in the data, with a failure to respond before the deadline on 3.53% of congruent trials, 4.30% of neutral trials, and 4.53% of incongruent trials.

Table 8
Results of the mixed effects linear regression for predicting reaction time in Experiment 3.

Fixed effect	*B*	β (95% CI)	*SE*	*t*	*p*
Intercept	874.325		10.467	83.527	< .001***
Word type (incongruent vs congruent)	26.166	0.06 (0.03, 0.09)	7.066	3.703	< .001***
Word type (neutral vs congruent)	16.246	0.04 (0.01, 0.07)	7.007	2.319	.021*
Frequency	−13.985	−0.07 (−0.10, −0.03)	3.663	−3.818	< .001***
Number of letters	17.596	0.08 (0.02, 0.15)	6.954	2.530	.012*
Number of Phonemes	3.426	0.02 (−0.03, 0.06)	5.134	0.667	.505
Orthographic Levenshtein Distance	−5.660	−0.03 (−0.09, 0,04)	6.757	−0.838	.403
Concreteness	2.114	0.01 (−0.02, 0.04)	3.048	0.693	.489
Valence	0.701	0.00 (−0.03, 0.04)	3.494	0.200	.841
Valence extremity	−0.593	−0.00 (−0.04, 0.03)	3.698	−0.163	.873
Arousal	−18.433	−0.09 (−0.12, −0.06)	3.157	−5.839	< .001***
Random effect	*SD*				
Word Intercept	38.739				
Participant Intercept	96.008				

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

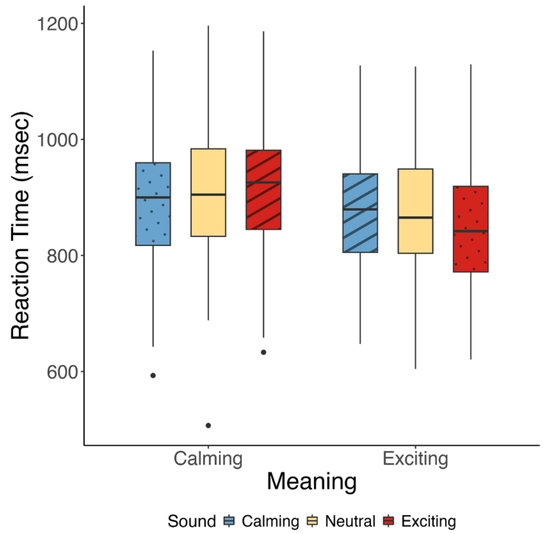

Fig. 5. Average reaction time by meaning and sound type in Experiment 3. *Note.* Dotted bars are congruent pairings, lined bars are incongruent pairings. The plot shows averages by subject.

a more exciting meaning ($b = 0.32, p < .001$). A supplementary analysis using incongruent words as the reference level found that their recall was not any different than neutral words ($b = 0.01, p = 0.95$).

In a supplementary analysis, we ran a version of the model that also included an interaction between word type and each word's meaning score, to examine whether congruency might affect exciting or calming words differently. This model found an interaction between the comparison of incongruent and congruent words, and word meaning ($p = .002$). In order to follow up on this interaction, we examined the effects of word type separately at high (SD = 2) and low (SD = −2) arousal. This revealed that congruent calming words were better remembered than incongruent calming words ($b = 1.12, p = .001$), but congruent exciting words were remembered *worse* than incongruent exciting words ($b = -0.68, p = .034$). A more straightforward interpretation of this pattern is that words with a calming sound were better remembered than words with exciting sounds.

Finally, we ran separate exploratory models that included an interaction between word type and scores on the verbal subscale of the IRQ or participant's Emotionality. There were no significant interactions (p's $>$ .33). The exploratory three-way interaction between word type, meaning score and Emotionality revealed a significant three-way interaction (congruent vs. neutral; $p = .043$). We followed this up using estimated marginal means and pairwise comparisons of high (SD = 2) and low values (SD = −2) of meaning score and Emotionality. This revealed an effect of emotional iconicity for calming words for those low in Emotionality ($b = 1.21, p = .01$), while incongruence for calming words had an effect for those high in Emotionality ($b = -1.02, p = .04$). Given the exploratory nature of this analysis and the high p-values we are hesitant to interpret this any further.

Discussion

The results of Experiment 3 replicate those of Experiment 2 on the semantic decision task with a much larger sample. Responses to iconic words were once again facilitated. We found evidence of both facilitation (in reaction time) and interference (in response accuracy). In addition, we observed that reaction time depended on a participant's level of Emotionality. In particular, for words with exciting meanings, iconicity (relative to neutral words) was facilitative for those high in Emotionality. Conversely, for words with calming meanings, congruent words were more facilitative for those low in Emotionality. In addition, regardless of word meaning, iconicity (relative to incongruent words) was facilitative for those low in Emotionality. We will return to a more fulsome interpretation of these interactions in the General Discussion.

Although we found a pattern in which iconicity was facilitative for the recall of calming words, a more straightforward interpretation of this

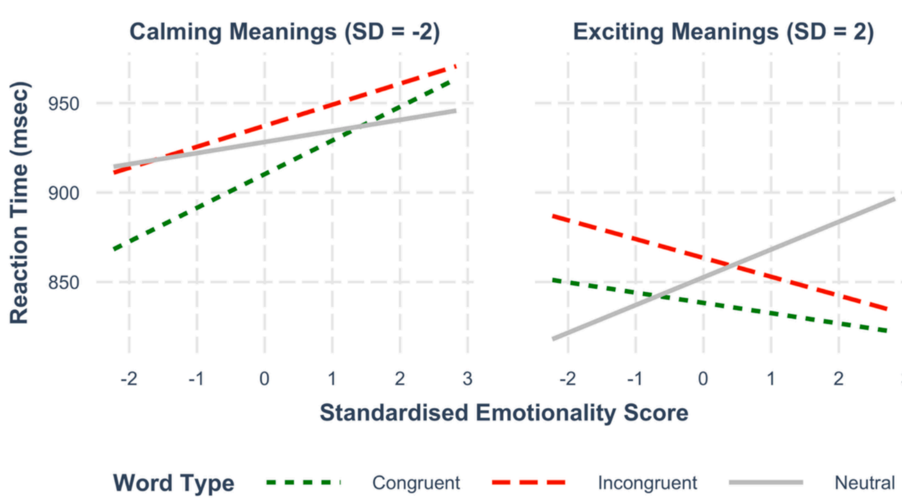

Fig. 6. Interaction between emotionality, word type and meaning score in the prediction of reaction time. *Note.* Estimated marginal means plotted.

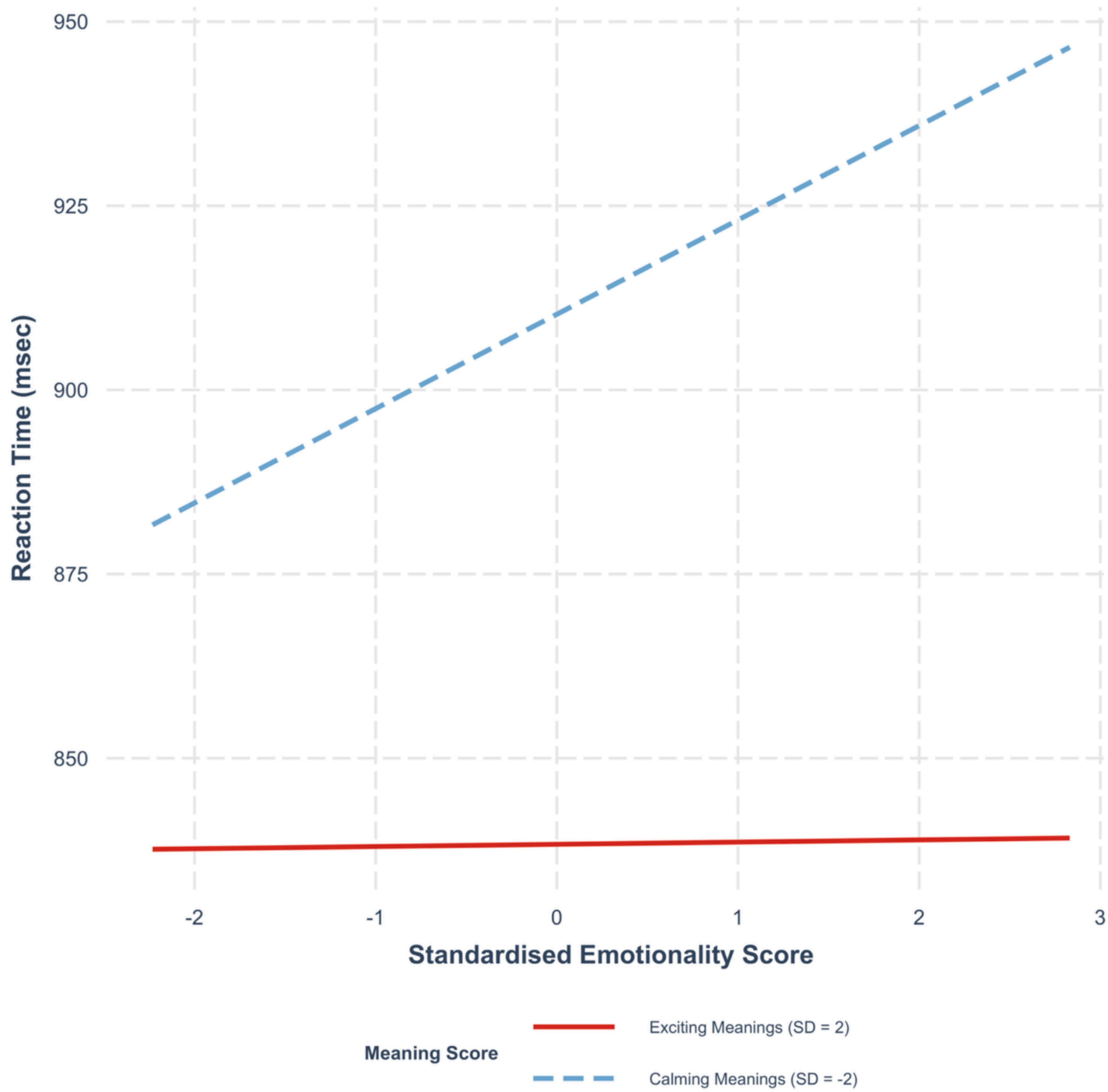

Fig. 7. Interaction between emotionality and meaning score in the prediction of reaction time. *Note.* Estimated marginal means plotted.

pattern was that words with calming sounds facilitated recall. Thus, the prudent interpretation is that Experiment 3 did not find evidence of iconicity affecting recall. In order to investigate the relationship between arousal iconicity and recall more fully, we next made use of an existing megastudy dataset. This also allowed us to examine whether iconicity would affect recall when words were not first categorized based on their emotion.

Recall megastudy analysis

To examine if the effects of emotional iconicity on recall generalize beyond our particular wordlists, and without a preceding emotional categorization task, we examined emotional iconicity in a recall memory megastudy dataset. In particular, we used the Penn Electrophysiology of Encoding and Retrieval Study (*PEERS*; available at: http://memory.psych.upenn.edu/Penn_Electrophysiology_of_Encoding_and_Retrieval_Study). The particular datasets that we analyzed were processed and organized by Madan (2021). The PEERS dataset includes data from several experiments in which participants were presented with lists of 16 items, followed by a 75 s recall interval. During encoding, participants were asked to make a size judgment, an animacy judgment, or were not given an encoding task (for more details, see Healey & Kahana, 2014; Lohnas & Kahana, 2013; Long et al., 2015). Here we collapsed across all encoding conditions and analyzed the overall recall probability for each word.

We first identified words that would have been categorized into each of the three categories from Experiment 2. There were 138 such words: 40 congruent, 54 neutral, and 44 incongruent. We then ran a multiple regression with probability of recall as the dependent variable, and the same predictors as in previous analyses. This revealed that congruent words were more likely to be recalled than incongruent words ($b = -0.05$, $p = .002$) or neutral words ($b = -0.04$, $p = .01$).[7] Words were also more likely to be recalled if they were more concrete ($b = 0.01$, $p = .03$) and had an extreme valence (i.e., extremely positive or negative; $b = 0.02$, $p = .03$). See Table 11. A supplementary analysis using incongruent words as the reference level found that their recall was not any different than neutral words ($b = 0.01$, $p = .39$). Supplementary analyses also indicated that these effects did not interact with word meaning score (p's $> .53$).

[7] These patterns remained after removing the one taboo word “breast”.

Table 9
Results of the mixed effects logistic regression for predicting accuracy in Experiment 3.

Fixed effect	*B*	exp(*B*)	ΔPercentage	*SE*	Wald's *z*	*p*
Intercept	1.315	3.725		0.092	14.336	< .001***
Word type (incongruent vs congruent)	−0.434	0.648	−8.115	0.119	−3.642	< .001***
Word type (neutral vs congruent)	−0.168	0.845	−2.936	0.115	−1.456	.146
Frequency	0.228	1.256	4.226	0.059	3.829	< .001***
Number of letters	−0.011	0.989	−0.201	0.113	−0.095	.924
Number of phonemes	−0.012	0.988	−0.224	0.084	−0.144	.885
Orthographic levenshtein distance	0.112	1.119	2.085	0.109	1.024	.306
Concreteness	−0.055	0.947	−1.014	0.049	−1.112	.266
Valence	−0.07	0.933	−1.305	0.057	−1.242	.214
Valence extremity	−0.071	0.932	−1.306	0.059	−1.176	.239
Arousal	−0.041	0.960	−0.759	0.052	−0.804	.422
Random effect	*SD*	*r*				
Word Intercept	0.680					
Participant Intercept	0.437					
Word Type (Incongruent) Slope	0.312	−0.442				
Word Type (Neutral) Slope	0.028	−0.472		.848		

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

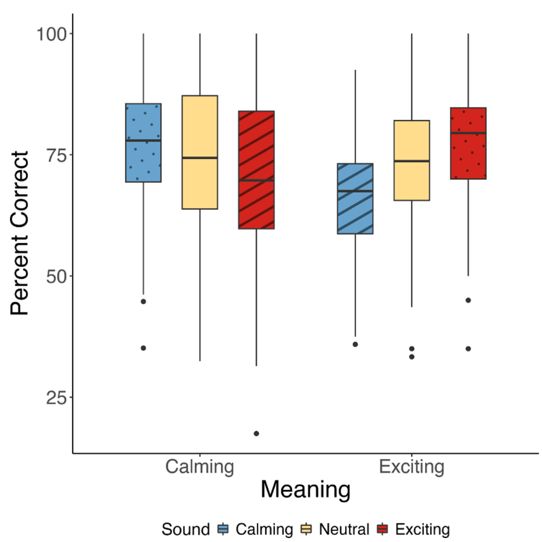

Fig. 8. Average percent correct by meaning and sound type in Experiment 3. *Note.* Dotted bars are congruent pairings, lined bars are incongruent pairings. The plot shows averages by subject.

In order to make use of the large number of words in the megastudy (i.e., not just words at the extremes), we calculated a derived measure of iconicity (henceforth *derived iconicity*) for each word (see Sidhu et al., 2021). In particular, we multiplied standardized meaning and sound scores. This results in positive ratings when both scores are either positive or negative (i.e., congruent), and negative scores when scores are

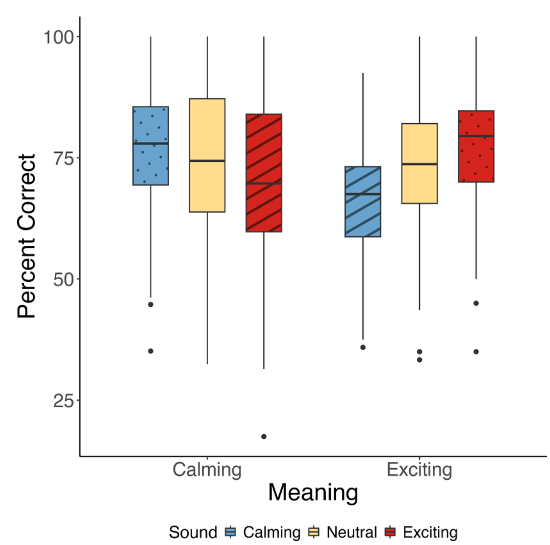

Fig. 9. Average percent recalled by meaning and sound type in Experiment 3. *Note.* Dotted bars are congruent pairings, lined bars are incongruent pairings. The plot shows averages by subject.

Table 10
Results of mixed effects logistic regression model predicting recall accuracy in Experiment 3.

Fixed effect	*B*	exp(*B*)	ΔPercentage	*SE*	Wald's *z*	*p*
Intercept	−3.166	0.042		0.115	−27.564	< .001***
Word type (incongruent vs congruent)	−0.184	0.832	−0.723	0.151	−1.221	.222
Word type (neutral vs congruent)	−0.174	0.840	−0.6865	0.149	−1.160	.246
Frequency	0.201	1.223	0.7727	0.077	2.592	.01**
Number of letters	−0.387	0.679	−1.4873	0.151	−2.565	.01*
Number of phonemes	−0.020	0.980	−0.0783	0.112	−0.182	.856
Orthographic levenshtein distance	0.416	1.516	1.597	0.145	2.859	.004**
Concreteness	0.141	1.151	0.5403	0.065	2.178	.029*
Valence	0.052	1.053	0.2017	0.073	0.715	.474
Valence extremity	0.063	1.065	0.2411	0.077	0.811	.417
Arousal	0.317	1.373	1.2175	0.066	4.770	< .001***
Random effect	*SD*					
Word intercept	0.799					
Participant intercept	0.466					

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

opposite in sign (i.e., incongruent). This approach has the added appeal of leading to larger values when either the sound or meaning are extreme and thus have more potential for iconicity. This approach captures iconicity along the entire spectrum. We included derived iconicity as the predictor of interest, along with previously mentioned control variables. These were available for 1,551 items. This revealed a significant effect of iconicity such that words that were higher in iconicity were more likely to be recalled ($b = 0.01$, $p < .001$). Words were also more likely to be recalled if they were more frequent ($b = 0.01$, $p < .001$), contained fewer phonemes ($b = -0.01$, $p = .04$), were more orthographically distinct (i.e., larger OLD; $b = 0.02$, $p < .001$), had a more extreme valence ($b = 0.01$, $p < .001$) and were more emotionally arousing ($b = 0.01$, $p < .001$). See Table 12 and Fig. 10. A supplementary analysis indicated that iconicity did not interact with word meaning score ($p = .26$).

General discussion

Iconicity refers to a perceived resemblance between form and meaning. Our particular interest here was in emotional iconicity. Certain phonemes are perceived as more exciting or calming and when they appear in words with an exciting or calming meaning, respectively (e.g., "chaotic", "lily"), they can be considered iconic. We examined whether such words would be processed faster and more accurately, and have a higher chance of being recalled.

The first major contribution of this work was a thorough characterization of phoneme-arousal associations (i.e., arousal sound symbolism). We found that voiceless stops, sibilants, short vowels and the high-front vowel /i/ were associated with excitement. Conversely nasal sonorants and long back vowels were associated with calmness. It is notable that these patterns replicate those that Aryani et al. (2018) found with German speakers, using a slightly different paradigm. While English and German are related (both Germanic languages) this nevertheless speaks to the generalizability of arousal iconicity. We did not find any evidence of these associations existing in the English lexicon when examining the entire range of arousal.

The phoneme-arousal associations we observed might arise from properties of the sounds themselves. For instance, it is notable that nasal sonorants consist of lower frequency sounds than both voiceless stops and sibilants, which consist of higher frequency sounds (Lacey et al., 2020). Back vowels also consist of a lower second formant, while the high-front vowel /i/ has a higher second format (Kawahara, 2021). This is consistent with a general association between high pitch and perceived states of arousal. That is, humans have a tendency to associate higher pitch vocalizations with higher states of arousal (i.e., "the pitch rule"; see Schwartz et al., 2024) in both human and non-human vocalizations alike. This may arise from the fact that a higher state of arousal does indeed raise the pitch of vocalizations. Sievers et al. (2019) have even suggested that higher frequency serves as a multi-sensory code for high emotional arousal, whether occurring auditorily or visually (e.g., in the contours of a shape; see also Aryani et al., 2020). Beyond pitch, Aryani et al. (2018) speculated that the association between short vowels and excitement may be due to the tendency to take quicker breaths when one is excited. Thus, phoneme-arousal associations may be rooted in the acoustic properties of vocalizations in heightened states of arousal.

In Experiments 2 and 3, we found that iconically congruent words were processed faster and more accurately on a semantic decision task (i.e., categorize each word as exciting or calming) as compared to incongruent words. While several studies have demonstrated that onomatopoeia are processed faster (Meteyard & Davies, 2020; Sidhu et al., 2020), the evidence for effects of crossmodal iconicity in language processing has been mixed (Aryani & Jacobs, 2018; cf. Sidhu & Pexman, 2022). The present results replicate Aryani and Jacobs (2018; with German stimuli) in an English-speaking population and English words. While English and German are closely related languages, Schmidtke and Conrad (2025) noted the consistent spelling-sound rules in German made it an ideal language for testing effects of iconicity with visually presented items. The present results are notable for extending emotional iconicity effects with visually presented stimuli to English, a language with less consistent spelling-sound rules (see Goswami et al., 2005). In addition, the results bolster the claim that emotional iconicity can affect language processing. They also demonstrate that associations that emerge in nonword rating tasks can affect the processing of real language, helping bridge the gap between sound symbolism studies and language processing more generally.

In the Introduction we discussed two potential mechanisms by which the processing of emotionally iconic words may be facilitated. The first was that congruence between sound and meaning could affect performance by facilitating the retrieval of a word's meaning (i.e., in semantic processing). In particular, Meteyard et al. (2015; see also Sidhu & Pexman, 2022) speculated that iconic words may have special links between sound and meaning, allowing facilitated retrieval of a word's semantics. One way of conceptualizing this is as more direct links between phonology and semantics, when articulatory or acoustic properties of phonemes also comprise an emotional experience. This captures the sense in which phonological features *are* semantic features (for a recent proposal involving Semantic Pointer Architecture, see Schmidtke, 2025). The second was that the sound of a word may facilitate/interfere with making a judgment about its emotion (see Aryani & Jacobs, 2018). As characterized by Aryani and Jacobs, cues from the sound of a word may be incorporated into decisions about the semantic content of a word, either facilitating or interfering with the decision. An extreme version of this mechanism is that the sound of a word may simply facilitate a decision of "exciting" or "calm" on a semantic decision task. Note that these interpretations localize the effect at the level of decision making, rather than semantic processing. The present results don't allow us to conclusively identify the mechanism at play here, but we offer speculation in the sections that follow. Future studies might use a task with a different decision category to help adjudicate between these possibilities.

Table 11
Results of the linear regression predicting recall of congruent, neutral, and incongruent words in the PEERS dataset.

Predictor	*B*	β (95% CI)	*SE*	*t*	*p*
Intercept	0.632		0.011	58.251	< .001***
Word type (incongruent vs. congruent)	−0.049	−0.32 (−0.53, −0.12)	0.016	−3.163	.002**
Word type (neutral vs. congruent)	−0.037	−0.26 (−0.45, −0.06)	0.014	−2.627	.01*
Frequency	0.008	0.11 (−0.07, 0.30)	0.007	1.206	.230
Number of letters	0.010	0.14, (−0.31, 0.59)	0.016	0.604	.547
Number of phonemes	−0.006	−0.08 (−0.40, 0.24)	0.011	−0.522	.603
Orthographic levenshtein distance	0.010	0.15 (−0.24, 0.54)	0.014	0.751	.454
Concreteness	0.013	0.19 (0.01, 0.36)	0.006	2.145	.034*
Valence	−0.004	−0.05 (−0.23, 0.12)	0.006	−0.590	.556
Valence extremity	0.017	0.24 (0.02, 0.45)	0.008	2.213	.029*
Arousal	0.011	0.15 (−0.05, 0.35)	0.007	1.475	.143

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

Table 12
Results of the linear regression predicting recall of words in the PEERS dataset.

Predictor	*b*	β (95% CI)	*SE*	*t*	*p*
Intercept	0.609		0.002	350.128	< .001***
Iconicity	0.007	0.01 (0.05, 0.15)	0.002	4.048	< .001***
Frequency	0.012	0.17 (0.11, 0.22)	0.002	6.253	< .001***
Number of letters	−0.001	−0.02 (−0.15, 0.11)	0.005	−0.302	.763
Number of phonemes	−0.008	−0.11 (−0.22, 0.00)	0.004	−2.040	.042*
Orthographic levenshtein distance	0.024	0.33 (0.22, 0.44)	0.004	6.072	< .001***
Concreteness	−0.001	−0.01 (−0.06, 0.04)	0.002	−0.549	.583
Valence	−0.001	−0.02, (−0.07, 0.04)	0.002	−0.588	.557
Valence extremity	0.009	0.13 (0.07, 0.18)	0.002	4.705	< .001***
Arousal	0.007	0.1 (0.05, 0.15)	0.002	3.901	< .001***

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

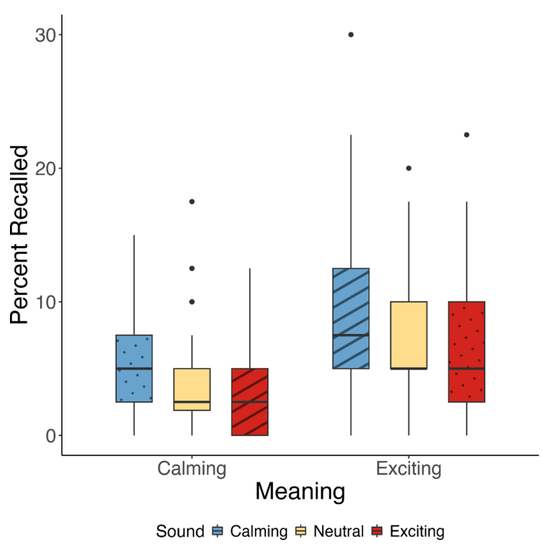

Fig. 10. Recall probability by derived iconicity. *Note.* Line is the estimated marginal effect of derived iconicity at mean levels of other variables in the model.

It is important to note that the recognition memory megastudy we analyzed did not involve making a decision about emotion during encoding. Further, because we examined recall rather than recognition in the mega study and Experiments 2 and 3, the effects we observed in memory cannot be attributed to facilitated decisions. This is one piece of evidence that emotional iconicity effects can extend beyond decision making.

The inclusion of neutral words allowed us insight into whether the effects were due to facilitation or interference. Interestingly, we observed facilitation in reaction time, and interference in response accuracy. In other words, iconic congruence sped up responses while incongruence led participants to the incorrect answer. The finding that iconic congruence facilitated response time while incongruence did not show the opposite pattern is consistent with an interpretation that localizes the effect in semantics. An effect at the level of decision making would have predicted a symmetrical effect. An effect at the level of processing makes no such prediction. While congruent words are predicted to have special links between sound and meaning, there is no expectation for incongruent to have *worse* links.

Previous studies which have found different patterns in reaction time and accuracy have speculated that different processes contributed to either (e.g., Kane & Engle, 2003). We might make the same speculation here. While faster responses to congruent words may reflect facilitated retrieval of their meaning, more inaccurate responses to incongruent items could reflect a decisional effect. It is worth noting that a supplementary analysis combining data from Experiments 2 and 3 found errors to incongruent items were faster than to neutral items ($b = 12.13$, $p = .038$). This may suggest participants incorporating sound cues (present in incongruent items but not neutral ones) into their decision-making process, leading to more confident, and incorrect, responses.

Individual differences

We also hoped to gain insight into whether emotional cues in word sound have an effect by eliciting emotional experiences in participants. This is orthogonal to the two mechanisms discussed above—evidence in favour of experienced emotion would not necessarily rule out either mechanism. Note that our main goal was to detect main effects of congruency, and thus interactions involving individual differences were likely underpowered.[8] Nevertheless, we included several individual difference measures relating to the experience of emotion. Our results were mixed. In Experiment 2 we did not observe any evidence of state anxiety nor Generalized Anxiety Disorder scores interacting with effects of iconicity. In Experiment 3 we did observe that a participant's Emotionality (i.e., the extent to which an individual tends to feel emotions such as fear, anxiety and empathy) moderated the effects of iconicity. These patterns were nuanced and complex. It seemed that iconic facilitation (i.e., congruent vs. neutral words) for words with calm meanings benefited processing time for those lower in Emotionality, while iconicity in exciting meanings benefited those higher in Emotionality. We might speculate that iconicity facilitates processing more so when it reflects an individual's general tendency towards emotion and their typical affective state (e.g., high arousal iconicity facilitated processing for those with a tendency towards strong emotions). We also discovered that incongruency in exciting words was interfering for those low in Emotionality. Perhaps for these words, individuals low in Emotionality rely more so on sound cues and are thus more likely to be led astray. Of course, these findings should be further explored in a more well powered follow up study that is better able to detect interactions with individual differences. At the moment, we cannot say whether effects of emotional iconicity arise due to emotional experiences elicited by language sounds. It would also be ideal to pair this task with a measure that can detect emotional experiences on a trial by trial basis (e.g., measures of pupil dilation or galvanic skin response).

Apart from iconicity, we also observed that those higher in Emotionality showed a larger reaction time benefit for more exciting meanings. This seemed to be primarily driven by slower responses to words with calming meanings, as a person's Emotionality increased. When examining the Big Five personality factors, Haro et al. (2024) found that those higher in Openness showed a larger effect of arousal. They did not find that any other personality factors had an effect. The present work is the first that we are aware of to demonstrate an association between Emotionality and arousal. Apart from these interactions, the general facilitation for exciting words is consistent with some previous work in lexical decision (e.g., Estes & Adelman, 2008;

[8] Indeed, a posthoc power analysis using the approach outlined in Kumle et al. (2021) suggested that our power to detect the one interaction that was observed in Experiment 3 (i.e., Word Type x Arousal x Emotionality) was only .72.

Haro et al., 2024). However, it should be noted that the effects of arousal in lexical decision are inconsistent, with some studies finding faster responses to calm words (Kuperman et al., 2014).

Effects of iconicity

It is worth contrasting these results with those of Sidhu and Pexman (2022) who found no effects of size iconicity on a semantic decision task. The link between phonemes and size may be less direct in nature than that between phonemes and arousal. For example, the link between /i/ and smallness is typically explained in one of two ways (see Sidhu & Pexman, 2018). One is that the small amount of open space in the mouth during articulation is connected with small size in the world through analogy. The other is that this association is an internalization of a statistical co-occurrence in the world in which smaller things resonate at a higher frequency. These may both be less direct than the association we have studied here, which is believed to result from a more general communicative association between sound and arousal (e.g., the pitch rule). Emotional iconicity may be uniquely positioned to affect fast paced word processing tasks, in contrast to other kinds of iconicity.

A worthwhile consideration is whether the phenomenon described so far is truly an instance of iconicity, or if it would be better characterized as indexicality. Whereas iconic signals resemble the things they stand for, indexical signals are "caused or affected by what they stand for" (p.4, Bateman, 2018). Iconicity and indexicality are both instances of non-arbitrariness. If certain acoustic properties are caused by feelings of excitement (i.e., if a person speaks with high pitch when excited), then these phoneme associations might rather be considered indexes. While these taxonomical issues are beyond the scope of this paper, if iconic and indexical relationships lead to different behavioral outcomes (e.g., differing effects observed with size vs. emotion), then there may be practical value in such considerations.

We also examined the possibility that these results could arise from patterns in the English language (i.e., systematicity). For example, if it were the case that exciting words in English tended to contain the sound /k/, an internalization of this pattern may have been what facilitated responses to congruent words (see Adelman et al., 2018; de Zubicaray et al., 2023; De Zubicaray et al., 2024; Reilly et al., 2012). However, there was no evidence that any phoneme's presence was predictive of a word's average arousal rating, when examining the entire range of arousal. This is somewhat surprising, as systematicity has been previously demonstrated for shape iconicity (Sidhu et al., 2021), size iconicity (Winter & Perlman, 2021) and texture iconicity (Winter, Sóskuthy, et al., 2022; for a review of iconic-systematicity see Sidhu, 2025). A key difference may have been that these studies looked for systematicity in specific kinds of words (e.g., shape iconicity in object nouns),[9] while here we looked at the lexicon in general, or at particular parts of speech. Future work might look for arousal systematicity among action words, for example. It is also worth noting that there is a great deal of variability in ratings of an item's arousal compared to, for example, valence (as noted by de Zubicaray et al., 2023). This would also make it more difficult to detect predictors of arousal. Another key difference may be that size or shape are defining features of a word's meaning. On the other hand, apart from words for emotional experience, this is not true for arousal. While "bomb" is extremely emotionally arousing, it is not part of its definition. This might moderate the extent to which types iconicity emerge in the lexicon. Finally, it is important to note that other studies have found evidence of arousal systematicity in the lexicon, when examined at the level of acoustic properties (Aryani et al., 2018) or sub-syllabic units (Schmidtke & Conrad, 2018) in German.

A novel contribution of this work is the demonstration that emotional iconicity has an effect in memory. Iconic words were more likely to be recalled in Experiment 2. However, this effect did not emerge in Experiment 3 after removing taboo items. Nevertheless, we observed a benefit in recall for emotionally iconic words in mega study dataset of over 1,500 items. This latter finding is notable because, unlike in Experiments 2 and 3, participants in the mega study were not specifically directed to process words in terms of their emotion. There was also no evidence that the effect of iconicity was moderated by word meaning in the mega study dataset.

At first glance this effect may seem to be driven by the well replicated finding that more emotionally arousing material is better remembered (e.g., Cortese & Khanna, 2022; Kensinger & Corkin, 2004). However, congruent words in the mega study were remembered better even when their meanings were calming. Instead, these results may be interpreted in terms of the richness of a word's encoding. Lau et al. (2018; see also Hargreaves et al., 2012) note that a word's semantic properties (e.g., the number of associated features) can provide an encoding affordance, leading to a memory trace that is more easily retrieved. They also note that this can make a word's trace more distinct, increasing the chances of it being correctly recalled. It could be that iconicity functions in this way. When a word's form enriches its emotional content, this may enrich its encoding. While admittedly speculative at this point, this is a promising area for future research, and a novel effect of iconicity. While previous work has shown effects of iconicity in recognition memory (Sidhu et al., 2023), this is the first demonstration of effects of iconicity in recall. This is also an effect of emotional iconicity which cannot simply be due to priming a decision category.

From a broader perspective, these results add to a growing appreciation that non-arbitrariness is an important factor in language (Dingemanse et al., 2015; Murgiano et al., 2021; Perniss et al., 2010). The form of a word can have a special imitative relationship with its meaning, with implications for language processing. Importantly, these results suggest that effects of iconicity go beyond imitations of sound (i.e., onomatopoeia) to crossmodal instances in which sound imitates emotion. This highlights the multimodal nature of language.

CRediT authorship contribution statement

David M. Sidhu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Pragati Parajuli:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Emiko J. Muraki:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in the form of a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (DGECR-2023-00428), and an Insight Development Grant from the Social Sciences and Humanities Research Council of Canada (430-2023-00418), to DMS. The authors thank Jeremy Armstrong, Jenna Brown, Sarah Goulding, Tim Hendrikx, Ellie Hollmann, Hayden Lacasse, Jason Scully, and Emily Thomas for their help collecting data; and Abby Traina for her help checking IPA transcriptions of words.

[9] There is evidence that the specific group of words tested matters. Sidhu et al. (2021) found no evidence of size systematicity in among object words, Winter and Perlman (2021) did among adjectives.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2026.104769

Data availability

Data and analysis code for all experiments can be found here: https://osf.io/ptfqy/.

References

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

Aryani, A., & Jacobs, A. (2018). Affective congruence between sound and meaning of words facilitates semantic decision. *Behavioral Sciences, 8*(6), 56. https://doi.org/10.3390/bs8060056

Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why "piss" is ruder than "pee"? The role of sound in affective meaning making. *PLoS One, 13*(6), Article e0198430. https://doi.org/10.1371/journal.pone.0198430

Aryani, A., Hsu, C. T., & Jacobs, A. M. (2019). Affective iconic words benefit from additional sound–meaning integration in the left amygdala. *Human brain mapping, 40* (18), 5289–5300.

Aryani, A., Isbilen, E. S., & Christiansen, M. H. (2020). Affective arousal links sound to meaning. *Psychological Science, 31*(8), 978–986. https://doi.org/10.1177/0956797620927967

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Bateman, J. A. (2018). Peircean semiotics and multimodality: Towards a new synthesis. *Multimodal Communication, 7*(1). https://doi.org/10.1515/mc-2017-0021

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389–411. https://doi.org/10.1037/met0000159

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

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

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

Conrad, M., Ullrich, S., Schmidtke, D., & Kotz, S. A. (2022). ERPs reveal an iconic relation between sublexical phonology and affective meaning. *Cognition, 226*, Article 105182. https://doi.org/10.1016/j.cognition.2022.105182

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

Cui, X., Damian, M. F., & Qu, Q. (2025). Co-activation of phonological and orthographic codes in various modalities of language processing: A systematic and meta-analytic review. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-025-02718-1

de Zubicaray, G. I., McMahon, K. L., Arciuli, J., Kearney, E., & Guenther, F. H. (2023). Emotion from the sound of a word: Statistical relationships between surface form and valence of English words influence lexical access and memory. *Journal of Experimental Psychology: General, 152*(12), 3566–3593. https://doi.org/10.1037/xge0001477

de Zubicaray, G. I., Arciuli, J., Guenther, F. H., McMahon, K. L., & Kearney, E. (2024). Non-arbitrary mappings between size and sound of English words: Form typicality effects during lexical access and memory. *Quarterly Journal of Experimental Psychology (2006), 77*(5), 943–963. https://doi.org/10.1177/17470218231184940

De Zubicaray, G. I., Kearney, E., Guenther, F., McMahon, K. L., & Arciuli, J. (2024). Statistical relationships between surface form and sensory meanings of English words influence lexical processing. *Journal of Experimental Psychology: Human Perception and Performance, 50*(7), 723–739. https://doi.org/10.1037/xhp0001209

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

Dingemanse, M., Perlman, M., & Perniss, P. (2020). Construals of iconicity: Experimental approaches to form–meaning resemblances in language. *Language and Cognition, 12* (1), 1–14. https://doi.org/10.1017/langcog.2019.48

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

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*, 1–22. https://doi.org/10.18637/jss.v033.i01.

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews, 68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Gallace, A., Boschin, E., & Spence, C. (2011). On the taste of "Bouba" and "Kiki": An exploration of word–food associations in neurologically normal participants. *Cognitive Neuroscience, 2*(1), 34–46. https://doi.org/10.1080/17588928.2010.516820

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

Goswami, U., Ziegler, J. C., & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: A comparison of English and German. *Journal of Experimental Child Psychology, Linguistic Constraints on Literacy Development, 92*(4), 345–365. https://doi.org/10.1016/j.jecp.2005.06.002

Hargreaves, I. S., Pexman, P. M., Johnson, J. S., & Zdrazilova, L. (2012). Richer concepts are better remembered: Number of features effects in free recall. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00073

Haro, J., Hinojosa, J. A., & Ferré, P. (2024). The role of individual differences in emotional word recognition: Insights from a large-scale lexical decision study. *Behavior Research Methods, 56*(8), 8501–8520. https://doi.org/10.3758/s13428-024-02488-z

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General, 143*(2), 575–596. https://doi.org/10.1037/a0033715

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), Article 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *The American Journal of Psychology, 121*(1), 83–103. https://doi.org/10.2307/20445445

Johansson, N., Anikin, A., & Aseyev, N. (2020). Color sound symbolism in natural languages. *Language and Cognition, 12*(1), 56–83. https://doi.org/10.1017/langcog.2019.35

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

Kawahara, S. (2021). *Phonetic bases of sound symbolism: A review* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fzvsu.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences, 101*(9), 3310–3315. https://doi.org/10.1073/pnas.0306408101.

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

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. Article 1 *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-05965-y.

Köhler, W. (1929). *Gestalt Psychology*. Liverlight.

Körner, A., & Rummer, R. (2022). Articulation contributes to valence sound symbolism. *Journal of Experimental Psychology: General, 151*, 1107–1114. https://doi.org/10.1037/xge0001124"

Kumle, L., Võ, M.-L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

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

Lacey, S., Jamal, Y., List, S. M., McCormick, K., Sathian, K., & Nygaard, L. C. (2020). Stimulus parameters underlying sound-symbolic mapping of auditory pseudowords to visual shapes. *Cognitive Science, 44*(9), Article e12883. https://doi.org/10.1111/cogs.12883

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

Lee, K., & Ashton, M. C. (2008). The HEXACO personality factors in the indigenous personality lexicons of English and 11 other languages. *Journal of Personality, 76*(5), 1001–1054. https://doi.org/10.1111/j.1467-6494.2008.00512.x

Lei, A., Willems, R. M., & Eekhof, L. S. (2023). Emotions, fast and slow: Processing of emotion words is affected by individual differences in need for affect and narrative absorption. *Cognition and Emotion, 37*(5), 997–1005. https://doi.org/10.1080/02699931.2023.2216445

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1943–1946. https://doi.org/10.1037/a0033669

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin & Review, 22*(5), 1328–1333. https://doi.org/10.3758/s13423-014-0791-2

Madan, C. R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability? *Psychonomic Bulletin & Review, 28*(2), 583–595. https://doi.org/10.3758/s13423-020-01820-w

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, Article 104092. https://doi.org/10.1016/j.jml.2020.104092

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia, Special Issue: Semantic Cognition, 76*, 264–275. https://doi.org/10.1016/j.neuropsychologia.2015.01.042

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. *Journal of Cognition, 4*(1), 38. https://doi.org/10.5334/joc.113

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech, 64*(1), 52–72. https://doi.org/10.1177/0023830920914339

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1*. https://doi.org/10.3389/fpsyg.2010.00227

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS One, 10*(9), Article e0137147. https://doi.org/10.1371/journal.pone.0137147

Peschard, V., Gilboa-Schechtman, E., & Philippot, P. (2017). Selective attention to emotional prosody in social anxiety: A dichotic listening study. *Cognition and Emotion, 31*(8), 1749–1756. https://doi.org/10.1080/02699931.2016.1261012

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual Word Recognition Volume 2: Meaning and Context, Individuals, and Development* (pp. 24–43). Psychology Press.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing [Computer software].

Reilly, J., Westbury, C., Kean, J., & Peelle, J. E. (2012). Arbitrary symbolism in natural language revisited: When word forms carry meaning. *PLoS One, 7*(8), Article e42286. https://doi.org/10.1371/journal.pone.0042286

Roebuck, H., & Lupyan, G. (2020). The internal representations questionnaire: Measuring modes of thinking. *Behavior Research Methods, 52*(5), 2053–2070. https://doi.org/10.3758/s13428-020-01354-y

Rummer, R., Schweppe, J., Schlegelmilch, R., & Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion, 14*(2), 246–250. https://doi.org/10.1037/a0035752

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12* (3), 225–239. https://doi.org/10.1037/h0070931

Scaltritti, M., Greatti, E., & Sulpizio, S. (2025). Decisional components of motor responses are not related to online response control: Evidence from lexical decision and speed-accuracy tradeoff manipulations. *Memory & Cognition, 53*(3), 911–925. https://doi.org/10.3758/s13421-024-01619-3

Schmidtke, D. (2025). *Echoes of Emotion: Effects of Affective Iconicity on Language Processing* [PhD]. University of Kassel.

Schmidtke, D., & Conrad, M. (2018). Effects of affective phonological iconicity in online language processing: Evidence from a letter search task. *Journal of Experimental Psychology: General, 147*(10), 1544–1552. https://doi.org/10.1037/xge0000499

Schmidtke, D., & Conrad, M. (2025). The role of valence and arousal for phonological iconicity in the lexicon of German: A cross-validation study using pseudoword ratings. *Cognition and Emotion, 39*(7), 1493–1514. https://doi.org/10.1080/02699931.2024.2353775

Schmidtke, D., Körner, A., Glim, S., & Rummer, R. (2025). Valence sound symbolism facilitates classification of vowels and emotional facial expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 51*(4), 661–667. https://doi.org/10.1037/xlm0001389

Schwartz, J. W., Pierson, K. H., & Reece, A. K. (2024). Pitch affects human (*Homo sapiens*) perception of emotional arousal from diverse animal calls. *Journal of Comparative Psychology, 138*(2), 80–87. https://doi.org/10.1037/com0000366

Sidhu, D. M. (2025). Sound symbolism in the lexicon: A review of iconic-systematicity. *Language and Linguistics Compass, 19*(1), Article e70006. https://doi.org/10.1111/lnc3.70006

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

Sidhu, D. M., & Pexman, P. M. (2022). Is a boat bigger than a ship? Null results in the investigation of vowel sound symbolism on size judgements in real language. *Quarterly Journal of Experimental Psychology*. , Article 17470218221078299. https://doi.org/10.1177/17470218221078299

Sidhu, D. M., Deschamps, K., Bourdage, J. S., & Pexman, P. M. (2019). Does the name say it all? Investigating phoneme-personality sound symbolism in first names. *Journal of Experimental Psychology: General, 148*(9), 1595–1614. https://doi.org/10.1037/xge0000662

Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition, 12*(1), 164–181. https://doi.org/10.1017/langcog.2019.36

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review, 28*(4), 1390–1398. https://doi.org/10.3758/s13423-021-01883-3

Sidhu, D. M., Williamson, J., Slavova, V., & Pexman, P. M. (2022). An investigation of iconic language development in four datasets. *Journal of Child Language, 49*(2), 382–396. https://doi.org/10.1017/S0305000921000040

Sidhu, D. M., Khachatoorian, N., & Vigliocco, G. (2023). Effects of iconicity in recognition memory. *Cognitive Science, 47*(11), Article e13382. https://doi.org/10.1111/cogs.13382

Sievers, B., Lee, C., Haslett, W., & Wheatley, T. (2019). A multi-sensory code for emotional arousal. *Proceedings. Biological Sciences, 286*(1906), Article 20190513. https://doi.org/10.1098/rspb.2019.0513

Simon, D., Becker, M., Mothes-Lasch, M., Miltner, W. H. R., & Straube, T. (2017). Loud and angry: Sound intensity modulates amygdala activation to angry voices in social anxiety disorder. *Social Cognitive and Affective Neuroscience, 12*(3), 409–416. https://doi.org/10.1093/scan/nsw131

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments* (Version R package version 0.16-1) [Computer software].

Spielberger, C. D. (1983). *State-trait anxiety inventory for adults*.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*, 1092–1097.

Sydeman, S. (2018). State-trait anxiety inventory. In V. Zeigler-Hill, & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_950-1.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology, 64*(12), 2392–2404. https://doi.org/10.1080/17470218.2011.605898

Uner, O., & Roediger, H. L., III (2022). Do recall and recognition lead to different retrieval experiences? *The American Journal of Psychology, 135*(1), 33–43. https://doi.org/10.5406/19398298.135.1.03

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., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language, 99*, 122–150. https://doi.org/10.1016/j.jml.2017.09.006

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics, 6*(1), Article 1. https://doi.org/10.5334/gjgl.1646

Winter, B., Perlman, M., Perry, L. K., Lupyan, G., & Dingemanse, M. (2022). Iconicity ratings for 14,000+English words. *OSF*. https://osf.io/y3wtx/.

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-021-04311-7

Winter, B., Woodin, G., & Perlman, M. (2023). *Winter, Woodin & Perlman—Defining iconicity for the cognitive sciences*. OSF. https://doi.org/10.31219/osf.io/5e3rc.

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

Yu, C.-S.-P., 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

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735