Cognition 274 (2026) 106550

Contents lists available at ScienceDirect

Cognition

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

Full Length Article

The word “bed” looks like a bed: Orthographic iconicity in English

David M. Sidhu [*]

Department of Psychology, Carleton University, Canada

ARTICLE INFO

Keywords:
Iconicity
Sound symbolism
Systematicity

ABSTRACT

Iconicity refers to the sense of resemblance between aspects of a signal's form and aspects of its meaning. This has traditionally been studied in the sound of words in spoken language, and the visual and kinesthetic properties of signs in signed languages. Here I ask the novel question of whether iconicity can be perceived in the shapes of words in a phonographic script (i.e., one in which symbols code for sounds) by native speakers. I collected ratings of orthographic iconicity (i.e., the extent to which the written form of a word resembles its meaning) for 3,150 English words. Unlike phonological iconicity, words high in orthographic iconicity tended to refer to visual meanings, speaking to the validity of the ratings and the distinction between these two forms of iconicity. Consistent with phonological iconicity, these words tended to be marked (e.g., contain infrequent letters and complex onsets), suggesting a general feature of iconicity. Two instances of objectively quantified resemblance between form and meaning (i.e., between the size and shape of written words and their referents) contributed to orthographic iconicity. Crucially, orthographic iconicity facilitated responses in several mega study datasets, even after controlling for suite of lexical semantic variables. These results suggest that participants can perceive iconicity in the visual appearance of words, and that it has measurable psychological effects.

Iconicity refers to a sense of resemblance between the form of a signal and its meaning (Winter, 2025). When appearing in language, it is an instance of non-arbitrariness, a way in which the forms of words can have a special relationship to their meanings. To illustrate, the English onomatopoeic words “splash”, “woof” and “bang” all resemble the sounds to which they refer. Iconicity in language has traditionally been studied in two domains. In spoken language research has examined the link between the sounds of words and their meanings (Perry et al., 2015; Winter et al., 2024). In signed language research has examined a link between the visual and kinesthetic properties of signs and their meanings (Emmorey, 2014). Here I examine the novel question of whether iconicity can be perceived between the written forms of words and their meanings in a phonographic script (i.e., one in which symbols code for sounds), and if this can have any measurable impacts on language processing.

It is important to note that iconicity is a subjective phenomenon. Contrary to objective definitions of iconicity, a more recent perspective has defined iconicity as the relationship between an individual's construal of a form and their construal of its meaning (Occhino et al., 2017). This definition allows for iconicity to exist across modalities. For example, a word such as “cactus” might be perceived as iconic when an individual construes its form as sounding abrupt or spiky, which resembles aspects of its meaning. This also highlights the fact that iconicity need not be absolute. Only a part of a signal's form needs to resemble a part of its meaning.

Iconicity is orthogonal to systematicity, another kind of non-arbitrariness in language. Systematicity refers to largescale patterns in the forms of words in a language with a similar usage (Dingemanse et al., 2015). For example, there are patterns in the forms of abstract vs. concrete words in English (e.g., concrete words tend to be shorter; Reilly et al., 2012). Recently de Zubicaray et al. (2023; 2024; 2024; Kearney et al., 2025) demonstrated that the formal properties of words (e.g., length, presence of particular phoneme groups) are predictive of their size, associated sensory experience, valence and concreteness. Further, these formal properties in words can affect their processing. For example, not only are words for larger objects responded to faster on word processing tasks, but words whose forms are typical of larger objects are also processed faster (de Zubicaray, Arciuli, et al., 2024). Systematicity is orthogonal to iconicity in that systematic patterns need not involve a sense of resemblance, though they may (for a review of iconic-systematicity, see Sidhu, 2025). Nevertheless, Imai and Akita (2023) speculated that systematic patterns in language could create a sense of “naturalness” that leads to higher ratings of iconicity. Therefore, one factor I will attend to is the extent to which perceptions of orthographic

* Corresponding author.
E-mail address: david.sidhu@carleton.ca.

https://doi.org/10.1016/j.cognition.2026.106550
Received 23 June 2025; Received in revised form 1 April 2026; Accepted 8 April 2026
Available online 21 April 2026

iconicity arise from systematicity.

As mentioned, research on iconicity in spoken language has focused on the sounds of words. It has been quantified by having participants rate the extent to which words sound like their meanings (Hinojosa et al., 2021; Perry et al., 2015; Thompson et al., 2020; Winter et al., 2017, Winter et al., 2024). Winter et al. (2024) collected the largest set of English iconicity ratings, by having participants rate over 14,000 words on a scale from 1 ("Not iconic at all") to 7 ("Very iconic"). This has allowed an exploration of the kinds of words that tend to be high and low in iconicity. For instance, iconic words tend to be high in sensory experience (Sidhu & Pexman, 2018; Winter et al., 2017; Winter et al., 2024), particularly in the auditory modality (Perlman et al., 2018). This is likely because auditory experience is more easily depicted by auditory forms. This is an important point as it illustrates that properties of the form and target domains can constrain iconic mappings (see Dingemanse et al., 2015; Punselie et al., 2024). The forms of iconic words also tend to be *marked* in that they stand out from other words in the lexicon (Dingemanse & Thompson, 2020). For instance, they tend to contain less common letters (Winter et al., 2024). This markedness has been argued to help highlight their depictive nature (Winter et al., 2024).

Importantly, research has shown that iconicity in spoken language has measurable psychological effects. It is argued that the resemblance between form and meaning may facilitate word learning, and indeed analyses of age of acquisition norms have found that iconic words tend to be learned earlier (Perry et al., 2015; Sidhu et al., 2022; for a review see Nielsen & Dingemanse, 2021). There is also evidence that iconic words (vs. non-iconic words) are processed faster on a lexical decision task (i.e., a task on which participants categorize letter strings as real words or nonwords, as quickly as possible; Meteyard et al., 2015; Sidhu et al., 2020). The argument is that iconic words may have special links between sound and meaning, thus aiding in retrieval of their meaning.

There is also an extensive literature of research on the iconicity of sign languages (see Emmorey, 2014; Perniss & Vigliocco, 2014). Unlike spoken language, sign language consists of visual kinesthetic signals, providing a rich source for iconic mappings to visual and movement-based features in meaning. In a further example of the differences afforded by signals in different modalities, Perlman et al. (2018) showed that the concreteness of meanings was a predictor of iconicity for words in signed but not spoken languages. Note that concreteness ratings are typically highest for visible and tangible meanings (Brysbaert et al., 2014). Research has shown that processing of iconic British Sign Language signs are facilitated on a variety of tasks including picture-sign matching (i.e., a task on which participants respond whether a visually presented sign matches a previously presented image) and naming (Vinson et al., 2015). There is also evidence of iconicity facilitating the acquisition of signs, though this is moderated by other factors including the type of iconicity (e.g., whether an action or perceptual feature is being depicted; see Ortega, 2017).

1. Iconicity in written language

While many studies on iconicity in spoken language have been conducted with written materials, participants are directed to consider the sounds of words. Here I explore whether iconicity can be perceived in the written form of words themselves. Previous studies have demonstrated iconicity in logographic systems, in which a single symbol can refer to a word or morpheme. For example, Xiao and Treiman (2012) demonstrated that non-speakers were above chance at guessing the meanings of a set of Chinese characters.[1] This may be due to the fact that some Chinese characters began as icons: pictures of objects, before undergoing a process of simplification to become less pictorial.

However, my interest here is in a phonographic system, in which symbols convey sounds rather than meanings. There is some evidence that symbols in these writing systems may be iconic of the sounds to which they refer. For example, Turoman and Styles (2017) found that individuals were above chance at guessing which of two symbols in unfamiliar languages coded for /i/ vs. /u/. There has also been some evidence that the shapes of letters contribute to the *maluma/takete effect* (Köhler, 1929; also referred to as the *bouba/kiki effect* more recently, Ramachandran & Hubbard, 2001). This is the tendency to pair certain phonemes (e.g., /b/, /u/) with round shapes, and others (e.g., /k/, /i/) with spiky shapes. Ćwiek et al. (2022) demonstrated that this association transcends any particular writing system. They had participants pair the nonwords "bouba" (typically round-associated) and "kiki" (typically spiky-associated) with a round and spiky shape, and found that speakers of 17 out of 25 languages made typical pairings. Interestingly, there was a non-significant trend for this typical pairing to be more likely in speakers of a language using the Roman writing system, in which "bouba" is written with round letters and "kiki" with spiky letters.

2. The present ratings

In the present study I ask whether the orthography of a phonographic writing system can afford iconic links between form and meaning. I examine this in English, which uses the Roman script in which symbols convey sounds. Importantly, my question is whether native English speakers can subjectively perceive and rate iconic links between form and meaning in such a system. To my knowledge this is the first time that anyone has explored whether a phonographic written system can afford iconic links.

I will use several principles to determine whether orthographic iconicity is a meaningful construct. If it is possible to meaningfully perceive orthographic iconicity in a phonographic system then there should be consistency in ratings across participants, on par with ratings of other psycholinguistic constructs. There should also be consistency in terms of the kinds of items which receive high and low ratings, and this should be compatible with a rating of visual resemblance (e.g., we would expect words higher in orthographic iconicity to refer to more visual referents). Moreover, there should be meaningful differences in the predictors of orthographic and phonological iconicity, to the extent that these are separate constructs. It will also be important to demonstrate that orthographic iconicity ratings are not wholly determined by another type of form-meaning congruence, namely systematicity. Finally, to the extent that orthographic iconicity is a relevant construct beyond the context of a rating task, there should be observable effects on behaviour (e.g., on the time it takes to process a visually presented word). These principles motivate the analyses to follow.

I begin by gathering ratings of orthographic iconicity. I then conduct analyses to explore various sets of theoretically important predictors of orthographic iconicity: visual features, specific form-meaning resemblances, markedness, and systematicity. I also report complementary analyses predicting phonological iconicity, to examine whether these are constructs that can be meaningfully separated. Finally, I examine whether orthographic iconicity has any measurable psychological effects on word processing and acquisition.

3. Method

3.1. Ethics statement

This study was approved by the Research Ethics Board at Carleton University. All participants provided informed consent. As per the approved ethics submission, participant anonymity was safeguarded by not storing personally identifying information in the data.

3.2. Participants

A total of 444 native English speakers took part in the rating task. Of

[1] Though see Occhino et al. (2017) for an important distinction between transparency of signals and iconicity.

these, 345 were undergraduate students from Carleton University who participated in exchange for course credit, while 99 were recruited from Prolific and participated in exchange for 4 GBP (paid at a rate of 16 GBP for a 15-min task). All participants reported English fluency and normal or corrected-to-normal vision. In addition, to match the undergraduate sample, participants on Prolific were required to have at least a high school education, reside in Canada or the United States, and be between the ages of 18 and 65. After data cleaning (see Results) a total of 384 participants ($M_{Age} = 23.86$, $SD_{Age} = 10.60$; 292 women, 84 men, four non-binary, one gender non-conforming, one gender fluid, two did not provide their gender) contributed to the final rating set. This left an average of 12.14 ratings per item, with 88% being rated by at least 10 raters (Motamedi et al., 2019).

3.3. Materials

The stimulus set consisted of 3,150 items. These were chosen because they had available ratings on phonological iconicity (Winter et al., 2024), as well as a range of other variables of interest. See Table 1 for a list of all variables examined, and the number of items for which these were available. Dimensions will be described in more detail as they appear in analyses.

3.4. Procedure

The experiment was run on Qualtrics. Participants received the following instructions, adapted from Winter et al. (2024):

> In this task, you will be rating some English words on their "iconicity."
> For each word, you will rate how well the appearance of the printed word itself (e.g., the shape of the letters) resembles its meaning.
> In other words, you are rating how well the appearance of the letters in a word resembles the meaning of that word.
> When a word looks like its meaning, it is iconic. When a word does not look like its meaning, it is not iconic.
> You will rate each word on a scale from 1 to 7. A rating of 1 indicates that the word is not at all iconic and does not at all look like what it means. 7 indicates that the word is high in iconicity and looks very much like what it means.
> Try to focus on the word meaning of the whole word, rather than decomposing it into parts. For example, when rating 'butterfly' think of the insect rather than "butter" and "fly," and rate how well the whole meaning relates to the sound of the whole word "butterfly."

Note that for the initial wave of data collection (42% of participants) there was a typo in the instructions which referred to the "sound" of the whole word butterfly (later corrected to "look"). Given subsequent analyses it does not appear that this affected most ratings. Participants also received a follow up question which clarified their task. That is, because of the unusual nature of the task, these instructions were followed by a multiple-choice question asking participants about the definition of "iconicity".

You will be rating words on their "iconicity". What does that mean?

- How much the visual appearance of the word resembles its meaning.
- How famous or well known the word is.
- How much a word affects you emotionally.
- How easy it was to picture the meaning of the word.

The answers appeared in a random order for each participant. An incorrect response returned participants to the instructions. If the participants answered incorrectly a second time their data were not used.

On each rating trial a participant was presented with a word along with a seven-point scale ranging from 1 ("Not at All Iconic") to 7 ("Very Iconic"). Each participant rated a random 100 items, in a random order. After these ratings there was an additional trial which explicitly asked the participant to choose a rating of 7. This was included as an attention check.

Participants then received a demographic questionnaire including questions about their age, gender, native language[2] and English fluency, an open-ended question regarding their understanding of iconicity, and whether there was any reason that their data should not be used.

Table 1
List of Variables Examined Along With Number Overlapping.

Dimension	N
Imageability (Graves et al., 2010)	2378
Visual Modality Strength (Lynott et al., 2020)	3150
Auditory Modality Strength (Lynott et al., 2020)	3150
Haptic Modality Strength (Lynott et al., 2020)	3150
Olfactory Modality Strength (Lynott et al., 2020)	3150
Gustatory Modality Strength (Lynott et al., 2020)	3150
Interoceptive Modality Strength (Lynott et al., 2020)	3150
Number of Letters	3150
Semantic Size (Scott et al., 2019)	1723
Letter Roundness (Porto et al., 2025)	3150
Shape Rating (Sidhu et al., 2021)	555
Unigram Frequency (Constrained; Medler & Binder, 2005)	3150
Unigram Frequency (Unconstrained)	3150
Bigram Frequency (Constrained)	3150
Bigram Frequency (Unconstrained)	3150
Spelling-Sound Consistency (Feedforward - Onset; Chee et al., 2020)	3150
Spelling-Sound Consistency (Feedforward - Nucleus)	3150
Spelling-Sound Consistency (Feedback - Onset)	3150
Spelling-Sound Consistency (Feedback - Nucleus)	3150
Orthographic Levenshtein Distance (Yarkoni et al., 2008)	3150
Systematicity (Monaghan et al., 2014)	1487
Sensory Experience Typicality (de Zubicaray, Arciuli, et al., 2024)	2407
Valence Typicality (de Zubicaray et al., 2023)	864
Size Typicality (de Zubicaray, Arciuli, et al., 2024)	1122
Concreteness Typicality (Kearney et al., 2025)	1068
Age of Acquisition (Kuperman et al., 2012)	3150
Age of Acquisition (Brysbaert & Biemiller, 2017)	3081

4. Results

All data and analysis code can be found at the following OSF repository: https://osf.io/msfy3.

4.1. Data cleaning

I removed participants who failed the attention check (3.61% of participants), answered the multiple-choice question about the task instructions incorrectly twice (4.74%), indicated that they answered randomly on the open-ended debriefing question (0%) or rated their English fluency lower than 5/5 (2.26%). Participants were also removed if their data indicated low effort responding. This was assessed in two ways. Participants who made the same response to each question were removed (0.45%). Participants were also removed if the correlation of their ratings with average ratings for each item in the group as a whole was lower than $r = 0.10$ (2.03%; Brysbaert et al., 2014). Note that these exclusion criteria were not mutually exclusive except that the correlations between participant ratings and item means were only run on participants who passed every other check. This left a total of 384 participants.

[2] Note that this was not included in the initial wave of data collection. With ethics approval, participants from that initial wave were contacted to ask about their native language. Of those contacted, 83 responded and 67 responded that they were native English speakers. These participants were included in the dataset.

4.2. Descriptives

The average rating was slightly above the midpoint of the scale ($M = 4.11$, $SD = 0.74$), see Fig. 1. Verbs were rated as the most iconic ($n = 437$, $M = 4.20$, $SD = 0.77$), followed by adjectives ($n = 200$, $M = 4.13$, $SD = 0.75$) and nouns ($n = 2483$, $M = 4.09$, $SD = 0.74$). The remaining 30 items were parts of speech with very few observations. This general pattern is consistent with ratings of phonological iconicity (Winter et al., 2024). The twenty words highest and lowest in orthographic iconicity are shown in Table 2.

Variation in orthographic iconicity ratings for each item differed across the range of the scale. In particular, ratings were more consistent at the extremes, and less consistent at the midpoint (see Fig. 2). Notably, this is what was observed for ratings of phonological iconicity (Winter et al., 2024) among other dimensions (e.g., concreteness; Brysbaert et al., 2014). Further, the average standard deviation in ratings of a given item on orthographic iconicity ($M_{SD} = 1.80$) is comparable to that of phonological iconicity ($M_{SD} = 1.82$). This suggests that participants were able to provide meaningful ratings, at least when using ratings of phonological iconicity as a benchmark. This was also comparable to other commonly used psycholinguistic ratings.[3] The average standard deviation among concreteness ratings ($N = 39{,}954$; Brysbaert et al., 2014) was 1.72, among valence ratings was 1.24, among arousal ratings in the same dataset was 1.73 ($N = 13{,}915$; Warriner et al., 2013), among imageability ratings ($N = 3000$; Cortese & Fugett, 2004) was 1.97 and among body object interaction ratings ($N = 9349$; Pexman et al., 2019) was 1.76.

4.3. Correlation between orthographic and phonological iconicity

Orthographic iconicity was significantly correlated with phonological iconicity ($r = 0.38$, $p < .001$), see Fig. 3. However, the two were not redundant. A correlation of 0.38 suggests that 14% of the variance in orthographic iconicity overlaps with phonological iconicity. This suggests that individuals were able to meaningfully rate orthographic iconicity as a construct distinct from phonological iconicity. See Appendix for zero-order correlations between orthographic and phonological iconicity and all variables in Table 1. See Online Supplementary Material (https://osf.io/msfy3) for a replication of the analyses run on phonological iconicity by Winter et al. (2024).

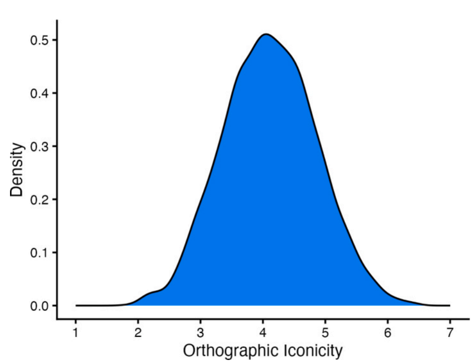

Fig. 1. Distribution of Orthographic Iconicity Ratings.

[3] Note that some ratings were not made on a seven-point scale (i.e., concreteness [5-point], valence [9-point], arousal [9-point]), and so these were rescaled to a seven-point scale. Of course, this is not a perfect comparison as a different rating scale would entail a different experience for the participant.

Table 2
Twenty Words Rated Highest and Lowest in Orthographic Iconicity (M; SD).

Highest Rated	Lowest Rated
bubble (6.45, 0.69)	patron (1.83, 0.75)
gulp (6.38, 1.04)	dinghy (2, 1.35)
crisp (6.38, 1.12)	morsel (2.09, 1.87)
zipper (6.38, 1.06)	revel (2.09, 1.04)
burp (6.3, 0.82)	paper (2.09, 1.14)
loop (6.23, 1.24)	herald (2.09, 1.04)
hoop (6.19, 0.91)	consul (2.1, 0.99)
sniff (6.18, 1.33)	premise (2.11, 1.17)
fluff (6.17, 1.19)	know (2.11, 1.76)
spark (6.09, 0.83)	deacon (2.14, 1.17)
sprinkle (6.09, 0.70)	thyroid (2.17, 1.19)
puppy (6.08, 1.08)	flier (2.17, 1.11)
baboon (6.08, 1.78)	caste (2.18, 1.17)
poof (6.08, 1.44)	permit (2.18, 0.75)
loose (6.08, 0.95)	sod (2.2, 1.62)
blizzard (6, 0.94)	convent (2.2, 1.32)
claw (6, 1.48)	tray (2.2, 2.17)
spoon (6, 1.41)	jackal (2.22, 1.39)
look (5.93, 1.22)	asphalt (2.25, 1.42)
love (5.93, 1.27)	coup (2.27, 1.68)

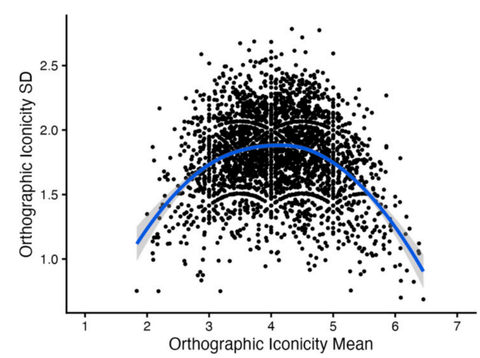

Fig. 2. Mean and Standard Deviation of Orthographic Iconicity Ratings. *Note.* Curve is a LOESS function.

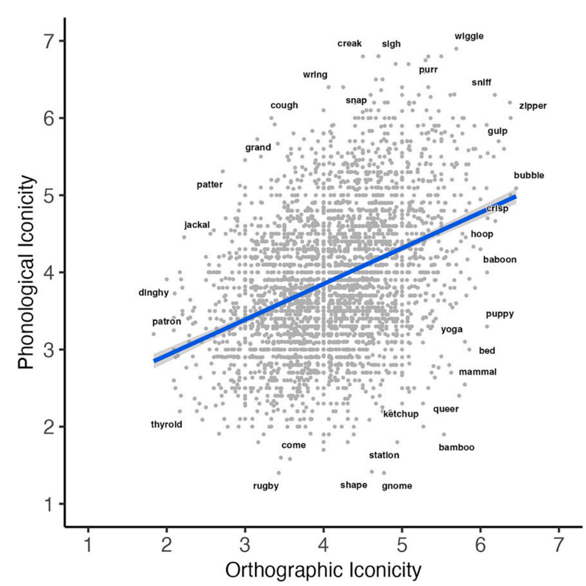

Fig. 3. Correlation Between Orthographic and Phonological Iconicity. *Note.* Phonological iconicity from Winter et al. (2024).

4.4. Predictors of orthographic iconicity

4.4.1. Visual features

4.4.1.1. Imageability. If participants were making meaningful ratings in terms of the extent to which word forms resemble their meanings, we would expect words with more visual features in their meaning to be better vehicles for iconicity. I examined this using a linear regression in which imageability (ratings of the ease with which one can generate a mental image of a referent; complied in Graves et al., 2010) as the predictor (n = 2,378). In this and all subsequently reported regression analyses, continuous predictors were always standardized. This revealed that imageability was a significant predictor of orthographic iconicity (b = 0.14, $p < .001$). Thus, for every one standard deviation increase in a word's imageability, it received a 0.14 higher rating of orthographic iconicity. Notably, this measure was not a significant predictor of phonological iconicity in the same items ($b = -0.01$, $p = .54$), supporting the possibility that participants are making distinct ratings in either case.

4.4.1.2. Perceptual strength. I also examined how well perceptual strength in six modalities (ratings of the extent to which a word is expereinced in a given modality; Lynott et al., 2020) predicted orthographic iconicity using a multiple linear regression model (n = 3,150). Visual ($b = 0.05$, $p < .001$), auditory ($b = 0.05$, $p = .001$), haptic (b = 0.10, p < .001), olfactory ($b = -0.01$, $p = .56$) interoceptive ($b = 0.06$, $p < .001$) and gustatory strength ($b = 0.04$, $p = .035$) all predicted orthographic iconicity, see Table 3. Notably, when these same predictors were entered in a model predicting phonological iconicity, visual strength was a significant negative predictor ($b = -0.07$, p < .001).

4.4.2. Resemblance

Examining iconicity at the level of ratings obscures the specific mappings that participants are attending to (see Thompson, Perlman, et al., 2020). While the specific relationships that individuals focus on when making ratings are likely idiosyncratic (both across items and individuals) there may be some salient mappings that are amenable to direct measure. One potential example is between the size of a word and its referent. I examined whether a simple potential iconic mapping would predict orthographic iconicity ratings: that between letter length and referent size. To that end, I computed a regression analysis including letter length, size ratings of a given concept (Scott et al., 2019) and their interaction (n = 1,723). There was a significant interaction between the two predictors ($b = 0.05$, $p = .008$), see Fig. 4. Follow up analyses using estimated slopes in the “emmeans” package in R (Lenth et al., 2022) indicated that for smaller concepts (standardized size rating = −1), shorter words were rated as more iconic ($b = -0.09$, p < .001). There was no effect of word length for larger concepts (standardized size rating = 1; $b = -0.005$, $p = .83$). Importantly, this interaction was not significant as a predictor of phonological iconicity ($p = .51$). This

Table 3
Results of Perceptual Strength in Five Modalities Predicting Orthographic and Phonological Iconicity.

	Prediction of Orthographic Iconicity		Prediction of Phonological Iconicity	
Predictor	*b*	*SE*	*b*	*SE*
Intercept	4.11***	0.01	3.90***	0.02
Visual Strength	0.05***	0.02	−0.06***	0.02
Auditory Strength	0.04**	0.01	0.07***	0.02
Haptic Strength	0.10***	0.02	0.18***	0.02
Olfactory Strength	−0.01	0.02	−0.08***	0.02
Gustatory Strength	0.04*	0.02	−0.00	0.02
Interoceptive Strength	0.06***	0.01	0.11***	0.02

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. $R^2_{\text{Orthographic}} = 0.03$. $R^2_{\text{Phonological}} = 0.06$. Analyses conducted on 3,150 items.

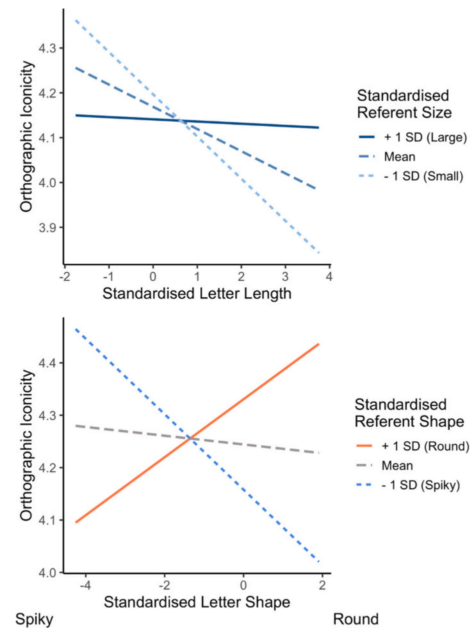

Fig. 4. Effects of Size and Shape Resemblance on Iconicity Ratings.
Note. Top plot shows the interaction between standardized letter length and referent size (Scott et al., 2019); analysis conducted on 1,723 items. Bottom plot shows the interaction between standardized letter shape and referent shape (Sidhu et al., 2022); analysis conducted on 555 items.

suggests that participants rated the orthography of a word as more iconic if its length matched the smaller size of its referent.

Another possible mapping is between the shapes of letters and the shapes of objects. Recently, Porto et al. (2025) found that participant ratings of letter shape correlated with first-order edge entropy, which captures the uniformity of edges in a shape (e.g., highest for “o”, “g” and “c”; lowest for “l”, “w” and “z”). For each word, I calculated the average first-order edge entropy (henceforth *letter shape*) of its letters. I computed a regression analysis including letter shape, shape ratings of a given object (Sidhu et al., 2022) and their interaction (n = 555). There was a significant interaction between the two predictors ($b = 0.06$, p = .04), see Fig. 4. Follow up analyses found no effect of letter shape for specifically spiky (standardized shape rating = −1; $b = -0.07$, $p = .14$) nor round shaped references (standardized shape rating = 1; $b = 0.06$, p = .14).

4.4.3. Markedness

4.4.3.1. Lexical variables. Previous work has demonstrated that iconic words tend to be structurally marked (Dingemanse, 2012; Winter et al., 2024). That is, they tend to consist of unusual forms which may aid their function as depictions by alerting individuals that they are performative in nature. To investigate a variety of formal predictors I computed a multiple regression analysis including letter and bigram frequency measures (both unconstrained, and constrained based on position in the word; Medler & Binder, 2005), letter length, and four measures of spelling-sound consistency (Chee et al., 2020; n = 3,150). These were feedforward and feedback consistency of each word's onset and nucleus. This quantifies how common a given spelling to sound mapping (feedforward) and sound to spelling mapping (feedback) is in English. Lastly, I included a measure of a word's orthographic neighbourhood density (edit distance between a word and its 20 closest orthographic neighbors; Yarkoni et al., 2008). Words high in orthographic iconicity had less frequent letters ($b = -0.06$, $p < .001$) and bigrams ($b = -0.06$, $p <. 001$) when measured without considering position (i.e., unconstrained). However, these were not significant predictors when considering

Table 4
Results of Formal Properties Predicting Orthographic and Phonological Iconicity.

	Prediction of Orthographic Iconicity		Prediction of Phonological Iconicity	
Predictor	*b*	*SE*	*b*	*SE*
(Intercept)	4.11***	0.01	3.90***	0.02
Unigram Frequency (Constrained)	−0.01	0.02	0.00	0.03
Unigram Frequency (Unconstrained)	−0.06***	0.02	−0.16***	0.02
Bigram Frequency (Constrained)	−0.00	0.02	0.01	0.02
Bigram Frequency (Unconstrained)	−0.06***	0.02	−0.06**	0.02
Number of Letters	−0.07**	0.03	0.05	0.03
Orthographic Levenshtein Distance	0.02	0.03	−0.18***	0.03
Spelling-Sound Consistency (Feedforward - Onset)	0.04**	0.01	0.07***	0.02
Spelling-Sound Consistency (Feedforward - Nucleus)	0.08***	0.01	0.12***	0.02
Spelling-Sound Consistency (Feedback - Onset)	−0.02	0.01	0.02	0.02
Spelling-Sound Consistency (Feedback - Nucleus)	−0.03*	0.01	0.02	0.02

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. $R^2_{\text{Orthographic}} = 0.05$. $R^2_{\text{Phonological}} = 0.12$. Analyses conducted on 3,150 items.

frequency at a particular position (p's > 0.75). Further, words with a more consistent spelling to sound relationship in their onset ($b = 0.04$, $p = .002$) and nucleus were also rated as more iconic ($b = 0.09$, $p < .001$). In addition, words with a more consistent sound to spelling relationship in their nucleus were also considered more iconic ($b = -0.03$, $p = .03$). See Table 4. The pattern was largely similar for phonological iconicity, reflecting the fact that markedness may be a feature common to iconicity regardless of modality.

4.4.3.2. Phonotactic complexity. I also considered the specific predictors of markedness that Dingemanse and Thompson (2020) showed were related to phonological iconicity: complex onsets (defined here as a word beginning with two or more consonant phonemes; e.g., “clip”), complex codas (a word ending with two or more consonant phonemes; e.g., “carp”), and the specific ending “-le”. As they did, I considered the words in the top 10 percentiles of orthographic iconicity as extremely iconic. I then ran a logistic regression analysis predicting orthographic iconicity extremity (i.e., a word belonging to top ten percent vs the rest of the distribution) using the presence of each of these cues as predictors ($n = 3{,}150$). This revealed that the words highest in orthographic iconicity were more likely to have complex onsets ($b = 0.36$, $p = .006$) and end in “-le” ($b = 0.67$, $p = .005$). Complex codas were not a significant predictor ($p = .22$). See Table 5.

4.4.4. Systematicity

While iconicity and systematicity are distinct concepts, it is possible that participants' ratings were affected by systematicity. That is, participants might be sensitive to largescale statistical relationships in the forms of words, and this may contribute to the sense that a word resembles its meaning. I analyzed this using several available measures of systematicity. The first was Monaghan et al.'s (2014) measure of systematicity which calculated the contribution of each word to the correlation between form and meaning in language. Computing a linear regression with systematicity as the predictor revealed that it was not a significant predictor of orthographic iconicity ($b = 0.01$, $p = .54$). A separate analysis revealed that it was also not a significant predictor of phonological iconicity ($b = 0.04$, $p = .13$). Analyses were conducted on 1,487 items.

In a series of studies, de Zubicaray et al. calculated the formal properties of English words that predict their sensory experience (2024), valence (2023), concreteness (Kearney et al., 2025) and size (2024). Using these properties, they were able to calculate the extent to which each word's form was typical of a meaning that is high or low on a given dimension. I computed four linear regressions which included a given measure of typicality, the relevant dimension and their interaction. If participants are sensitive to typicality of a form when rating orthographic iconicity, one would predict an interaction such that, for example, words referring to large sizes received higher ratings when their form is typical of a large object. The interaction was not significant in the prediction of sensory experience ($p = .48$; $n = 2{,}407$), valence ($p = .43$; $n = 864$), nor concreteness ($p = .27$; $n = 1{,}068$). The interaction was significant in the case of size ($b = 0.06$, p $= .005$; $n = 1{,}122$). Follow up analyses suggested that size typicality had no effect on orthographic iconicity for large objects (standardized size $= 1$; $b = -0.02$, $p = .45$). However, for small objects (standardized size $= -1$), those with forms more typical of a small object forms were rated as more iconic ($b = -0.14$, $p < .001$). Thus, this instance of systematicity may have contributed to iconicity ratings. For analyses of typicality and phonological iconicity see Online Supplementary Material (https://osf.io/msfy3).

Table 5
Results of Phonotactic Complexity Predicting a Word Belonging in Top Ten Percentiles of Orthographic and Phonological Iconicity.

	Prediction of Orthographic Iconicity		Prediction of Phonological Iconicity	
Predictor	*b*	*SE*	*b*	*SE*
(Intercept)	−2.29***	0.08	−2.64***	0.09
Complex Onset	0.36**	0.13	1.13***	0.12
Complex Coda	−0.19	0.16	0.27	0.15
Ends in “-le”	0.67**	0.24	1.27***	0.22

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. Analyses conducted on 3,150 items.

4.5. Interim summary

I examined several theoretically informative predictors of orthographic iconicity. In contrast to the prediction of phonological iconicity, high imageability and visual perceptual strength were both significant predictors of orthographic iconicity. There was also some evidence of resemblance between form and meaning (based on size and shape) affecting orthographic iconicity ratings. As in the prediction of phonological iconicity, words high in orthographic iconicity tended to have more marked forms (i.e., have less frequent letters and bigrams; contain complex onsets and the suffix “-le”). There was only one instance of systematicity predicting ratings, in words for small things with typically small forms being rated as more iconic.

4.6. Psychological effects of orthographic iconicity

Thus far the results suggest that participants can rate orthographic iconicity when asked to, and that there are patterns in the kinds of words considered to be iconic. I next examined the validity of these ratings by testing whether they predicted behaviour in several mega study datasets and norms. This would suggest that orthographic iconicity is relevant beyond a context in which individuals are explicitly asked to consider it.

4.6.1. Lexical decision

I made use of three available lexical decision datasets: the English lexical decision project (*ELP*; Balota et al., 2007), the British lexical decision project (*BLP*; Keuleers et al., 2012), and the English crowdsourcing project (*ECP*; Mandera et al., 2020). The ELP dataset was collected in person across six American universities, testing native English speakers. Participants were presented with 14 blocks of 250 trials, with the final block containing fewer trials, across two testing sessions. Participants were presented with letter strings to categorize as real words or meaningless nonwords by button press as quickly as possible. A trial began with three asterisks appearing in the center of the screen for 250 milliseconds (ms), followed by a 50 ms tone, and a 250 ms blank screen. The target word was presented in the location where the three asterisks had been. Participants received feedback if they were inaccurate, or were told that they were too slow if a response was not given within 4000 ms. There was a 1000 ms blank screen between trials.

The BLP was collected with a slightly different procedure. Participants came from one British university, and took part in person. Participants were presented with 57 blocks of 500 trials, followed by a final block of 230 trials. Blocks were completed over multiple sessions. Participants could take a break every 100 trials. As in the ELP, participants categorized letter strings as words or nonwords. A trial began with two horizontal lines presented onscreen for 500 ms. The target was then presented between these lines. Following a response there was a 500 ms blank screen between trials. There was no time limit for responses.

The ECP was conducted online, testing native English speakers. Participants were presented with 100 trials consisting of 70 words and 30 nonwords. They could participate as many times as they liked. Unlike the ELP and BLP, participants were asked to indicate if they knew each word or not (rather than whether a target was a word or nonword). They were told not to guess if they were not sure. They were also not encouraged to answer quickly.

The dependent variables in all subsequent mega study analyses was the average response time to each word (after standardizing response times within participants) and average response accuracy to each word. To rule out any possible confounding variables, I controlled for a suite of variables that previous work has shown predict lexical decision times (see Brysbaert et al., 2019). This included contextual diversity, word frequency (Brysbaert and New, 2009), prevalence (Brysbaert et al., 2019), self-report age of acquisition (Kuperman et al., 2012), number of letters, unconstrained bigram frequency (Medler and Binder, 2005), orthographic Levenshtein distance (Yarkoni et al., 2008), number of phonemes, number of syllables, feedforward spelling sound consistence of onset and nucleus (Chee et al., 2020), valence, valence extremity, arousal (Warriner et al., 2013), composite perceptual strength (Minkowski distance at m = 3; Lynott et al., 2020), sensory experience (Juhasz and Yap, 2013), imageability (Graves et al., 2010), number of semantic neighbors (Shaoul and Westbury, 2010), number of senses (Miller, 1995) and part of speech (Brysbaert et al., 2012). Because phonological iconicity has been shown to affect word processing (Meteyard et al., 2015; Sidhu et al., 2020) I controlled for it as well.

Orthographic iconicity predicted significantly faster reaction time in the ELP ($b = -0.02$, $p = .01$), BLP ($b = -0.03$, $p = .007$) and ECP ($b = -0.01$, $p < .001$). It also predicted significantly more accurate responses in the BLP ($b = 0.01$, $p < .001$) and ECP ($b = 0.002$, $p < .001$), but not the ELP ($b = 0.00$, $p = .92$). It is notable that phonological iconicity predicted lower accuracy in the ELP ($b = -0.003$, $p = .02$) and ECP ($b = -0.002$, $p = .048$), as well as slower responding in the ECP ($b = 0.006$, $p < .001$), while being a non-significant predictor in other cases (p's > 0.10). See Table 6 for reaction time results, and Table 7 for accuracy results.

4.6.2. Naming

I next examined the naming data in the ELP. The procedure was the same as the ELP lexical decision task, except that there were no nonwords, and the participants' task was to say a word out loud rather than categorize it as a word or nonword. Participants were presented with 12 blocks of 250 trials, with the final block containing more trials. This analysis removed bigram frequency and part of speech as controls (see Brysbaert et al., 2019). Orthographic iconicity was not a significant predictor of naming time ($p = .40$), it was however a significant predictor of naming accuracy ($b = 0.002$, $p = .006$). See Table 8.

4.6.3. Semantic decision

Finally, I examined the semantic decision data in the Calgary Semantic Decision project (Pexman et al., 2017). This task consisted of concrete and abstract words (based on previous ratings; Brysbaert et al., 2014). Participants took part in person and received four blocks of 250 items. They could take a break between blocks. The participant's task was to categorize a word as concrete or abstract as quickly as possible. A trial began with two horizontal lines presented onscreen for 500 ms. The target was then presented between these lines. Following a response there was a 500 ms blank screen between trials. Participants received feedback if they were inaccurate, or were told that they were too slow if a response was not given within 3000 ms. In the interest of comparison I included the same control variables as in the analyses of lexical decision.

Table 6
Resulting Models Predicting Reaction Time in Four Datasets.

	ELP RT		ECP RT		BLP RT		CSDP RT	
Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	−0.971***	0.022	−0.600***	0.003	−0.901***	0.023	−0.122***	0.036
Orthographic Iconicity	−0.023**	0.009	−0.007***	0.001	−0.025**	0.009	−0.013	0.014
Phonological Iconicity	0.007	0.009	0.006***	0.001	0.014	0.010	0.043**	0.015
Contextual Diversity	−0.438***	0.046	−0.046***	0.006	−0.476***	0.049	−0.167	0.092
Frequency	0.279***	0.045	0.025***	0.006	0.311***	0.048	0.039	0.090
Prevalence	−0.123***	0.009	−0.048***	0.001	−0.184***	0.010	−0.027	0.016
Age of Acquisition (Self-Report)	0.036**	0.012	0.003*	0.002	0.063***	0.013	0.078***	0.021
Number of Letters	0.067***	0.017	0.020***	0.002	0.080***	0.018	0.036	0.028
Average Bigram Frequency (Unconstrained)	0.012	0.008	0.000	0.001	−0.012	0.009	0.018	0.013
Orthographic Levenshtein Distance	0.041**	0.015	−0.008***	0.002	−0.016	0.016	0.005	0.024
Number of Phonemes	−0.010	0.013	−0.004*	0.002	−0.016	0.014	0.004	0.021
Number of Syllables	0.025*	0.012	−0.001	0.002	−0.010	0.013	−0.055**	0.020
Spelling-Sound Consistency (Feedforward - Onset)	−0.003	0.008	−0.002*	0.001	−0.007	0.009	−0.020	0.013
Spelling-Sound Consistency (Feedforward - Nucleus)	−0.002	0.008	0.000	0.001	−0.007	0.008	−0.005	0.012
Valence	−0.010	0.008	−0.004**	0.001	−0.008	0.009	0.009	0.013
Valence Extremity	0.010	0.009	−0.002*	0.001	−0.014	0.009	−0.007	0.015
Arousal	0.006	0.009	−0.001	0.001	−0.016	0.009	0.002	0.015
Concreteness	0.054***	0.014	0.010***	0.002	0.031*	0.015	−0.145***	0.026
Perceptual Strength	−0.016	0.010	−0.002	0.001	−0.026*	0.011	−0.025	0.017
Sensory Experience Rating	−0.009	0.011	−0.003*	0.001	0.022	0.011	−0.014	0.019
Imageability	−0.071***	0.014	−0.013***	0.002	−0.072***	0.016	−0.053	0.028
Semantic Neighbors	−0.086***	0.013	−0.007***	0.002	−0.080***	0.014	0.063**	0.020
Number of Senses	0.005	0.010	0.002	0.001	−0.010	0.011	0.046**	0.015

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. $R^2_{ELP} = 0.53$. $R^2_{ECP} = 0.71$. $R^2_{BLP} = 0.58$. $R^2{}_{CSDP} = 0.56$. The first three analyses conducted on 2,284 items; the last analysis conducted on 582 items. Part of speech coefficients omitted for readability. ELP = English Lexicon Project Lexical Decision Task (Balota et al., 2007). ECP = English Crowdsources Project Recognition Task (Mandera et al., 2020). BLP = British Lexicon Project Lexical Decision Task (Keuleers et al., 2012). CSDT = Calgary Semantic Decision Task (Pexman et al., 2017).

Table 7
Resulting Models Predicting Accuracy in Four Datasets.

	ELP Accuracy		ECP Accuracy		BLP Accuracy		CSDP Accuracy	
Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	0.944***	0.003	0.990***	0.001	0.952***	0.004	0.838***	0.014
Orthographic Iconicity	0.000	0.001	0.002***	0.000	0.006***	0.002	0.010	0.005
Phonological Iconicity	−0.003*	0.001	−0.001*	0.000	−0.003	0.002	−0.016**	0.006
Contextual Diversity	0.047***	0.007	0.004**	0.001	0.053***	0.008	0.121**	0.038
Frequency	−0.035***	0.007	−0.003*	0.001	−0.036***	0.008	−0.086*	0.037
Prevalence	0.042***	0.001	0.015***	0.000	0.053***	0.002	0.010	0.006
Age of Acquisition (Self-Report)	−0.005*	0.002	0.000	0.000	0.001	0.002	−0.008	0.008
Number of Letters	0.002	0.003	0.000	0.001	0.005	0.003	0.003	0.011
Average Bigram Frequency (Unconstrained)	−0.002	0.001	0.000	0.000	0.001	0.001	−0.009	0.005
Orthographic Levenshtein Distance	−0.007**	0.002	0.000	0.000	0.001	0.003	0.005	0.009
Number of Phonemes	0.006**	0.002	0.000	0.000	0.002	0.002	0.006	0.008
Number of Syllables	0.001	0.002	0.000	0.000	0.001	0.002	0.005	0.008
Spelling-Sound Consistency (Feedforward - Onset)	0.001	0.001	0.000	0.000	0.000	0.001	0.004	0.005
Spelling-Sound Consistency (Feedforward - Nucleus)	0.002	0.001	0.000	0.000	−0.001	0.001	−0.002	0.005
Valence	0.000	0.001	0.000	0.000	−0.001	0.002	−0.001	0.005
Valence Extremity	−0.001	0.001	0.000	0.000	−0.002	0.002	0.002	0.006
Arousal	0.000	0.001	0.000	0.000	0.001	0.002	−0.003	0.006
Concreteness	−0.014***	0.002	−0.001***	0.000	−0.007**	0.003	0.040***	0.010
Perceptual Strength	−0.001	0.002	−0.001	0.000	−0.002	0.002	−0.003	0.007
Sensory Experience Rating	0.000	0.002	−0.001	0.000	0.001	0.002	0.005	0.007
Imageability	0.020***	0.002	0.003***	0.000	0.017***	0.003	0.021	0.011
Semantic Neighbors	0.001	0.002	0.000	0.000	−0.003	0.002	−0.021**	0.008
Number of Senses	0.001	0.002	0.000	0.000	0.002	0.002	−0.008	0.006

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. $R^2_{\text{ELP}} = 0.55$. $R^2_{\text{ECP}} = 0.69$. $R^2_{\text{BLP}} = 0.55$. $R^2_{\text{CSDP}} = 0.36$. The first three analyses conducted on 2,284 items; the last analysis conducted on 582 items. Part of speech coefficients omitted for readability. ELP = English Lexicon Project Lexical Decision Task (Balota et al., 2007). ECP = English Crowdsources Project Recognition Task (Mandera et al., 2020). BLP = British Lexicon Project Lexical Decision Task (Keuleers et al., 2012). CSDT = Calgary Semantic Decision Task (Pexman et al., 2017).

Table 8
Resulting Models Predicting Reaction Time and Accuracy in ELP Naming Task.

	ELP Naming RT		ELP Naming Accuracy	
Predictor	b	SE	b	SE
Intercept	−0.884***	0.008	0.986***	0.001
Orthographic Iconicity	−0.008	0.009	0.002**	0.001
Phonological Iconicity	−0.007	0.010	0.001	0.001
Contextual Diversity	−0.099*	0.048	0.006	0.004
Frequency	0.021	0.046	−0.004	0.004
Prevalence	−0.066***	0.010	0.010***	0.001
Age of Acquisition (Self-Report)	0.026*	0.013	−0.003**	0.001
Number of Letters	0.184***	0.017	−0.005***	0.001
Orthographic Levenshtein Distance	0.030*	0.015	0.001	0.001
Number of Phonemes	0.019	0.014	0.003*	0.001
Number of Syllables	−0.084***	0.013	0.001	0.001
Spelling-Sound Consistency (Feedforward - Onset)	−0.079***	0.009	0.004***	0.001
Spelling-Sound Consistency (Feedforward - Nucleus)	−0.024**	0.008	0.003***	0.001
Valence	−0.007	0.009	−0.002**	0.001
Valence Extremity	−0.002	0.009	0.000	0.001
Arousal	0.005	0.009	−0.001	0.001
Concreteness	0.024	0.015	−0.002	0.001
Perceptual Strength	−0.019	0.011	0.000	0.001
Sensory Experience Rating	0.006	0.011	0.001	0.001
Imageability	−0.002	0.015	0.001	0.001
Semantic Neighbors	−0.053***	0.014	0.002	0.001
Number of Senses	0.020	0.010	0.000	0.001

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. $R^2_{\text{RT}} = 0.34$. $R^2_{\text{Accuracy}} = 0.19$. Analyses conducted on 2,284 items. ELP = English Lexicon Project Naming Task (Balota et al., 2007).

Notably this includes the significant predictors from Pexman et al. (i.e., frequency and sensory experience). Orthographic iconicity was not a significant predictor of reaction time ($b = -0.01, p = .36$) nor accuracy ($b = 0.01, p = .07$). Notably, phonological iconicity interfered with this task, slowing responses ($b = 0.04, p = .005$) and leading to less accurate responding ($b = -0.02, p = .009$). Lupyan and Winter (2018) previously speculated that this could be due to abstract iconic words seeming more concrete by evoking more specific meanings. See Table 6 for reaction time and Table 7 for accuracy results.

I also conducted separate versions of these analyses on only concrete or abstract items as different sets of variables will be relevant for either kind of item (Pexman et al., 2017). Orthographic iconicity was not a significant predictor in any of these analyses (p's > 0.08).

4.6.4. Age of acquisition

Age of acquisition has previously been included as a predictor of iconicity. I also examined whether orthographic iconicity predicted age of acquisition as a dependent variable, after controlling for phonological iconicity, child-directed speech frequency derived from the CHILDES database (MacWhinney, 2000),[4] length, orthographic Levenshtein distance (Yarkoni et al., 2008), concreteness (Brysbaert et al., 2014), familiarity (Brysbaert et al., 2014), imageability (Graves et al., 2010), valence and valence extremity (Warriner et al., 2013). I considered both retrospective age of acquisition (i.e., adult judgments of age at which an individual understood the word; Kuperman et al., 2012) and test based age of acquisition (i.e., the grade at which a majority of children could identify the correct word meaning in multiple choice; Brysbaert & Biemiller, 2017). I omitted words that did not appear in the CHILDES corpora from which child-directed speech frequency was derived (the iconicity results do not change if they are included). Orthographic

[4] These were calculated using the following North American English dialogues: Bates, Bernstein, Bliss, Bloom, Brown, Clark, Demetras, Gathercole, Hall, Higginson, Kuczaj, Peters, Post, Sachs, Snow, Suppes, Valian, VanHouten, VanKleeck, Warren. Each dialogue included an adult speaking to a child between the ages of 11 months and 11 years 8 months. Child-directed speech frequency was calculated as the number of times a lemma was spoken by an adult. A value of 1 was added to these values and then they were log-transformed. This is similar to previous derived frequencies (see Goodman et al., 2008). These were retrieved from the Child Language Data Exchange System (CHILDES; MacWhinney, 2000). Transcripts involving structured interviews or experiments were omitted. More information on these corpora can be found here: https://talkbank.org/childes/access/Eng-NA/.

Table 9
Resulting Models Predicting Age of Acquisition.

	Self-Report AoA		Test-Based AoA	
Predictor	*b*	*SE*	*b*	*SE*
Intercept	6.757***	0.027	6.843***	0.022
Orthographic Iconicity	−0.081**	0.030	−0.097***	0.024
Phonological Iconicity	−0.092**	0.030	−0.119***	0.025
CD Speech Frequency	−1.38***	0.03	−1.31***	0.03
Number of Letters	0.120*	0.049	0.131***	0.039
Orthographic Levenshtein Distance	−0.012	0.049	0.031	0.040
Concreteness	−0.045	0.050	0.021	0.040
Familiarity	−0.20***	0.03	−0.20***	0.02
Imageability	−0.464***	0.050	−0.519***	0.040
Valence	−0.063*	0.029	−0.074**	0.024
Valence Extremity	−0.140***	0.030	−0.121***	0.024

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. $R^2_{\text{Self-Report}} = 0.66$. $R^2_{\text{Test-Based}} = 0.74$. Self-report analysis conducted on 1,943; test-based analysis conducted on 1,923. AoA = age of acquisition. Self-report AoA (Kuperman et al., 2012) is in years, test-based AoA (Brysbaert & Biemiller, 2017) is in grade.

iconicity was a significant predictor of both retrospective age of acquisition ($b = -0.08$, $p = .008$) and test based age of acquisition (Brysbaert & Biemiller, 2017; $b = -0.10$, $p < .001$). See Table 9.

5. Discussion

I examined whether participants could meaningfully rate the iconicity of a phonographic script (i.e., orthographic iconicity). The results suggested that this was possible. The standard deviation in ratings was comparable to other semantic norms. Words high in orthographic iconicity also had certain features in common (e.g., were highly imageable) which suggested participants were meaningfully performing the rating task. Importantly, these words were also processed faster in three lexical decision mega study datasets, and more accurately in two out of the three datasets. They were also named more accurately and acquired earlier.

The findings were consistent with the possibility that participants were able to meaningfully rate the resemblance between the shape of a word and its meaning. Importantly, iconicity is based on a subjective sense of resemblance (Occhino et al., 2017). In some cases, this resemblance may be imagistic. For example, the word “blob” contains rounded letters and refers to something round. The word “bed” might resemble a bed viewed from the side. In other cases, this sense of resemblance may be more indirect in nature. For instance, the word “blizzard” contains zig-zagging letters which could evoke sharp gusts of wind. The word “echo” begins and ends with a rounded letter, perhaps analogically mapping onto a repeating sound. Importantly, iconic mappings do not require that the entire signal map onto the entire meaning (see Emmorey, 2014; Punselie et al., 2024). There only need to be aspects of form which map onto aspects of meaning.

Words high in orthographic iconicity tended to refer to more visual meanings. That is, they were high in imageability and referred to meanings experienced visually. This is consistent with participants rating the resemblance between the visual appearance of words and their referents. Notably, this was not the case for phonological iconicity, which had a non-significant relationship with imageability and a negative relationship with visual strength. This further establishes orthographic and phonological iconicity as separate constructs. This also highlights the fact the modality of a signal will constrain iconic mappings (see Dingemanse et al., 2015; Punselie et al., 2024). It is consistent with Perlman et al. (2018) who found differences in the kinds of sensory experience that predicted iconicity in spoken vs. signed languages. For instance, concreteness was a predictor of iconicity in ASL but not in English. The present ratings will allow researchers to compare iconicity across modalities within a single language.

When the resemblance between the visual appearance of words and their referents was quantified directly (i.e., via letter length or shape), there was evidence that this affected iconicity ratings. Due the subjective nature of iconicity, there have been questions as to whether ratings actually capture a sense of resemblance (for a discussion see Winter & Perlman, 2021). Showing that a direct quantification of resemblance predicts ratings suggests that they do capture a sense of resemblance. This is in line with research showing that nonwords rated as “round sounding” acoustically resemble the sounds made by round objects (Fort & Schwartz, 2022).

Analyses revealed that words high in both orthographic and phonological iconicity tended to be marked. For instance, they both contain less frequent letters, and were more likely to begin with complex onsets (e.g., “sp-”). This markedness might invite individuals to consider a given word as performative, thus helping arrive at an iconic construal (see Dingemanse & Thompson, 2020). Thus, markedness is not necessarily iconic in and of itself, but might prompt an individual to perceive the iconicity present in a given word. This is related to the notion of *coerced iconicity*: iconicity that is perceived when the depictive nature of a wordform invites one to think about its iconicity (Dingemanse, 2011). The fact that this was observed for both kinds of iconicity suggests that it is a general property of iconic forms.

The emergence of iconicity in language is an inherently difficult subject to address because of a lack of historical data.[5] However, one point worth making is to draw attention to a difference in the plausible origins of phonological and orthographic iconicity. Some examples of phonological iconicity stem from words coined as imitations (e.g., “meow”; for a discussion, see Akita & Imai, 2022). This is less likely for orthographic iconicity, as it is implausible to assume words were coined based on their appearance (apart from purposeful irregular spellings; e.g., “lite”).[6]

5.1. Alternate possibilities

Thus far the discussion has assumed that participants genuinely rated the iconic resemblance between word forms' visual appearance and their referents. However it is worth exploring other possible contributions to ratings. Another possibility is that participants were sensitive to an instance of systematicity or typicality. For example, Imai and Akita (2023) noted that examples of the phonestheme “gl-” (appearing often in words for light such as “glisten” or “gleam”) received very high iconicity ratings in Winter et al. (2024) despite lacking a clear iconic grounding. They speculated that this pattern might create a sense of “naturalness” that leads to an iconic construal. However, only one of the analyses I conducted (i.e., size typicality) found a relationship between measures of systematicity and orthographic iconicity. Perhaps the particular relevance of size typicality (vs. other instances of typicality) is that it involves some amount of iconic-systematicity (see Sidhu, 2025).

It is also possible that participants were making ratings on the basis of one or more lexical semantic variables. That is, perhaps participants developed a heuristic that, for example, words referring to imageable things with infrequent letters would receive a high rating. Or perhaps a sense of fluency in processing (e.g., spelling-sound consistency) was used as a proxy for iconicity. A final related possibility is that iconicity was facilitated—not determined alone—by certain lexical semantic variables. That is, participants may have more readily found iconic mappings for words with marked forms. Collecting qualitative data on the motivation for iconic ratings could help adjudicate between these possibilities.

[5] Though there have been excellent efforts to approach this question using analyses of reconstructed language roots (Winter et al., 2022), historical language change (Monaghan & Roberts, 2021), and iterated language learning tasks over successive “generations” of participants (Vinson et al., 2021).

[6] I thank an anonymous reviewer for this specific example.

5.2. Prediction of behaviour

It is especially notable that ratings of orthographic iconicity predicted responses to words in three lexical decision datasets, and in a naming task, even after controlling for a suite of relevant variables. This finding is important as it suggests that even though ratings of orthographic iconicity are subjective, they characterized a dimension which generalizes beyond the rating task and to a different set of participants. This shows that subjective perceptions of iconicity are a meaningful construct that affects language processing. Previous work has speculated that phonologically iconic words enjoy special links between sound and meaning, facilitating their retrieval. Those special links fit neatly when the sound of a word is imitating another sound meaning (e.g., "splash"), but I am wary of making such a claim about the physical resemblance between the shape of the word "bubbles" and the shapes of bubbles, for example. Nevertheless, it is an interesting possibility for future research to explore. Earlier I speculated that orthographic iconicity ratings could be caused by instances of systematicity or iconic-systematicity that were not fully captured in the analyses. Perhaps these facilitated retrieval (rather than orthographic iconicity itself). Further, if participants' ratings are indeed sensitive to some kind of fluency in the link between spelling and meaning, then this would also lend itself to the faster retrieval of word meanings.

The results in lexical decision also help further separate orthographic and phonological iconicity. Phonological iconicity was either not a significant predictor, or was a significant inhibitory predictor, in the word processing megastudies I examined. It is notable that these tasks involved orthographic processing, which may help explain why orthographic iconicity was the better predictor. It is also worth underscoring the null results for phonological iconicity in this large item set. Previous studies have demonstrated an effect of phonological iconicity on lexical decision tasks (Meteyard et al., 2015; Sidhu et al., 2020). However, the items in Meteyard et al. were largely onomatopoeia, and the analyses in Sidhu et al. suggested that effects of iconicity were most prominent in onomatopoeia. Thus, it may be that the facilitatory effects of phonological iconicity in lexical decision do not extend to the more general set of words analyzed here.

Both orthographic and phonological iconicity predicted earlier acquisition of words, whether measured through self-report or multiple choice questions at different grades. Indeed, both forms of iconicity were significant predictors when entered together into these models, suggesting they explained unique variance in both acquisition norms. This is a surprising result, as spoken language is acquired auditorily before children learn the written forms of words. It is beyond the scope of this paper to interrogate how this might take place. It may be that one of the other lexical semantic variables which predicted orthographic iconicity explains this result.

6. Conclusion

The main contribution of this work is to show that iconicity can be perceived in the orthography of a phonographic script. There are also meaningful patterns in the words that tend to be orthographically iconic. Orthographically iconic words tend to be higher in visual features which contrasts with phonological iconicity. Both orthographic and phonological iconicity tend to have marked forms, capturing the role of iconicity in depiction. There was also some evidence of perceptual resemblance between written form and meaning contributing to ratings of orthographic iconicity. Finally, orthographic iconicity has measurable psychological effects, facilitating processing on several tasks.

Iconicity is often presented as a foil to arbitrariness. However, this is not the case I am intending to make here. None of the preceding content would lead one to suspect that the shapes of English words are non-arbitrarily related to their meanings in general. Instead, the important takeaway is that any system which involves signals and meanings can afford a subjective sense of resemblance between the two. The present work reinforces that iconicity is a modality-general phenomenon, which can even be perceived in a phonographic writing system. It is predictable, with some mappings being perceived more consistently. Where it does appear, it can have measurable psychological effects. This work is a novel and unexpected demonstration of yet another way in which human communication is multimodal.

CRediT authorship contribution statement

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

Acknowledgements

The author thanks Daria Kievsky and Sofia Mah for their help editing several tables, Christen Potvin for her help with data management and quality checks, Rabten Tsering for computing letter shape scores for lowercase English letters, and Timothy Hendrikx for compiling child-directed speech frequency.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (DGECR-2023-00428) and Social Sciences and Humanities Research Council (430-2023-00418).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2026.106550.

Data availability

Available at https://osf.io/msfy3.

References

Akita, K., & Imai, M. (2022). The iconicity ring model for sound symbolism. In S. Lenninger, O. Fischer, C. Ljungberg, & E. Tabakowska (Eds.), *18. Iconicity in language and literature* (pp. 27–46). John Benjamins Publishing Company. https://doi.org/10.1075/ill.18.02aki.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods, 49*(4), 1520–1523. https://doi.org/10.3758/s13428-016-0811-4

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

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

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*, 991–997.

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

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

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384–387. https://doi.org/10.3758/BF03195585

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., … Winter, B. (2022). The *bouba/kiki* effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 377*(1841), 20200390. https://doi.org/10.1098/rstb.2020.0390

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. (2011). Ezra pound among the Mawu: Ideophones and iconicity in Siwu. In P. Michelucci, O. Fischer, & C. Ljungberg (Eds.), *10. Iconicity in language and* *literature* (pp. 39–54). John Benjamins Publishing Company. https://doi.org/10.1075/ill.10.03din.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Lang & Ling Compass, 6*(10), 654–672. https://doi.org/10.1002/lnc3.361

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., & Thompson, B. (2020). Playful iconicity: Structural markedness underlies the relation between funniness and iconicity. *Language and Cognition, 12*(1), 203–224. https://doi.org/10.1017/langcog.2019.49

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Fort, M., & Schwartz, J. L. (2022). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific Reports, 12*(1), 19172.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex, 20*(8), 1799–1815. https://doi.org/10.1093/cercor/bhp245

Hinojosa, J. A., Haro, J., Magallares, S., Duñabeitia, J. A., & Ferré, P. (2021). Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods, 53*(3), 1262–1275. https://doi.org/10.3758/s13428-020-01496-z

Imai, M., & Akita, K. (2023). The iconicity ring hypothesis bridges the gap between symbol grounding and linguistic relativity. *Topics in Cognitive Science, 15*(4), 676–682. https://doi.org/10.1111/tops.12671

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

Kearney, E., McMahon, K. L., Guenther, F., Arciuli, J., & de Zubicaray, G. I. (2025). Revisiting the concreteness effect: Non-arbitrary mappings between form and concreteness of English words influence lexical processing. *Cognition, 254*, Article 105972. https://doi.org/10.1016/j.cognition.2024.105972

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

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

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

Lenth, R. V., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). Emmeans: Estimated marginal means, aka least-squares means (version 1.8.2) [computer software] https://CRAN.R-project.org/package=emmeans.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society, B: Biological Sciences, 373*(1752), 20170137. https://doi.org/10.1098/rstb.2017.0137

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (Third Edition). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Medler, D. A., & Binder, J. R. (2005). *MCWord: An On-Line Orthographic Database of the English Language*. http://www.neuro.mcw.edu/mcword/.

Medler, D. A., & Binder, J. R. (2005). MCWord: An orthographic wordform database. https://www.neuro.mcw.edu/mcword/.

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 Special Issue: Semantic Cognition. *Neuropsychologia, 76*, 264–275. https://doi.org/10.1016/j.neuropsychologia.2015.01.042

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41. https://doi.org/10.1145/219717.219748

Monaghan, P., & Roberts, S. G. (2021). Iconicity and diachronic language change. *Cognitive Science, 45*(4), Article e12968. https://doi.org/10.1111/cogs.12968

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

Motamedi, Y., Little, H., Nielsen, A., & Sulik, J. (2019). The iconicity toolbox: empirical approaches to measuring iconicity. *Language and Cognition, 11*(2), 188–207.

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

Occhino, C., Anible, B., Wilkinson, E., & Morford, J. P. (2017). Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture, 16*(1), 100–126. https://doi.org/10.1075/gest.16.1.04occ

Ortega, G. (2017). Iconicity and sign lexical acquisition: A review. *Frontiers in Psychology, 8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01280.

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: A comparison between American sign language, British sign language, English, and Spanish. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01433

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 369*(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300

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

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: Concrete/abstract decision data for 10,000 English words. *Behavior Research Methods, 49*(2), 407–417. https://doi.org/10.3758/s13428-016-0720-6

Pexman, P. M., Muraki, E., Sidhu, D. M., Siakaluk, P. D., & Yap, M. J. (2019). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods, 51*(2), 453–466. https://doi.org/10.3758/s13428-018-1171-z

Porto, A., Huckle, N., Basalyga, A., Santiago, J., & Kranjec, A. (2025). Glyph norming: Human and computational measurements of shape angularity in writing systems. *Behavior Research Methods, 57*(6), 173. https://doi.org/10.3758/s13428-025-02682-7

Punselie, S., McLean, B., & Dingemanse, M. (2024). The anatomy of iconicity: Cumulative structural analogies underlie objective and subjective measures of iconicity. *Open Mind, 8*, 1191–1212. https://doi.org/10.1162/opmi_a_00162

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.

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

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

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*(2), 393–413. https://doi.org/10.3758/BRM.42.2.393

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

Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience, 33*(1), 25–31. https://doi.org/10.1080/23273798.2017.1358379

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

Thompson, A. L., Akita, K., & Do, Y. (2020). Iconicity ratings across the Japanese lexicon: A comparative study with English. *Linguistics Vanguard, 6*(1). https://doi.org/10.1515/lingvan-2019-0088

Thompson, B., Perlman, M., Lupyan, G., Sehyr, Z. S., & Emmorey, K. (2020). A data-driven approach to the semantics of iconicity in American sign language and English. *Language and Cognition, 12*(1), 182–202. https://doi.org/10.1017/langcog.2019.52

Turoman, N., & Styles, S. J. (2017). Glyph guessing for 'oo' and 'ee': Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *Royal Society Open Science, 4*(9), Article 170882. https://doi.org/10.1098/rsos.170882

Vinson, D., Jones, M., Sidhu, D. M., Lau-Zhu, A., Santiago, J., & Vigliocco, G. (2021). Iconicity emerges and is maintained in spoken language. *Journal of Experimental Psychology: General, 150*(11), 2293–2308. https://doi.org/10.1037/xge0001024

Vinson, D., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British sign language. *Journal of Memory and Language, 82*, 56–85. https://doi.org/10.1016/j.jml.2015.03.002

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

Winter, B. (2025). The size and shape of sound: The role of articulation and acoustics in iconicity and crossmodal correspondencesa. *The Journal of the Acoustical Society of America, 157*(4), 2636–2656. https://doi.org/10.1121/10.0036362

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2024). Iconicity ratings for 14,000+ English words. *Behavior Research Methods, 56*(3), 1640–1655. https://doi.org/10.3758/s13428-023-02112-6

Winter, B., & Perlman, M. (2021). Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard, 7*(1). https://doi.org/10.1515/lingvan-2020-0135

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies, 18*(3), 443–464. https://doi.org/10.1075/is.18.3.07win

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), Article 1. https://doi.org/10.1038/s41598-021-04311-7

Xiao, W., & Treiman, R. (2012). Iconicity of simple Chinese characters. *Behavior Research Methods, 44*(4), 954–960. https://doi.org/10.3758/s13428-012-0191-3

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond colthearts N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

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., 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