Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-021-01883-3

BRIEF REPORT

Sound symbolism shapes the English language: The maluma/takete effect in English nouns

David M. Sidhu[1] · Chris Westbury[2] · Geoff Hollis[3] · Penny M. Pexman[1]

Accepted: 13 January 2021

Abstract

Sound symbolism refers to associations between language sounds (i.e., phonemes) and perceptual and/or semantic features. One example is the maluma/takete effect: an association between certain phonemes (e.g., /m/, /u/) and roundness, and others (e.g., /k/, /ɪ/) and spikiness. While this association has been demonstrated in laboratory tasks with nonword stimuli, its presence in existing spoken language is unknown. Here we examined whether the maluma/takete effect is attested in English, across a broad sample of words. Best–worst judgments from 171 university students were used to quantify the shape of 1,757 objects, from spiky to round. We then examined whether the presence of certain phonemes in words predicted the shape of the objects to which they refer. We found evidence that phonemes associated with roundness are more common in words referring to round objects, and phonemes associated with spikiness are more common in words referring to spiky objects. This represents an instance of iconicity, and thus nonarbitrariness, in human language.

Keywords Sound symbolism · Iconicity · Maluma/takete effect · Bouba/kiki effect · Shape judgment

When shown the shapes in Fig. 1, roughly 90% of individuals worldwide (Styles & Gawne, 2017; see this paper also for exceptions) associate the round one with nonword labels like *maluma* and the spiky one with nonword labels like *takete.* This *maluma/takete effect* (Köhler, 1929) is one example of *sound symbolism*, associations between formal language sounds and perceptual and/or semantic properties.[1] These associations may derive from perceptuomotor analogies between language sounds and properties in other modalities (e.g., between the abrupt sounds in *takete* and the abrupt changes in outline of the spiky shape; see Sidhu & Pexman, 2018). The maluma/takete effect has been observed in speakers of different languages (Styles & Gawne, 2017) and across ages (for a review see Fort et al., 2018). It has been demonstrated using both behavioural tasks (e.g., Nielsen & Rendall, 2011) and neuroimaging (e.g., Asano et al., 2015).

The maluma/takete effect generalizes to several categories of phonemes. The most robust associations are between sonorants (e.g., /l/, /m/, /n/) and round shapes, and voiceless stops (e.g., /p/, /t/, /k/) and spiky shapes (e.g., McCormick, Kim, List, & Nygaard, 2015). Some have also found associations between voiced stops (e.g., /b/, /d/, /g/; most commonly /b/) and round shapes (e.g., McCormick et al., 2015). In addition, studies have shown associations of rounded back vowels (e.g., /u/ as in *boot*) with round shapes, and unrounded front vowels (e.g., /i/ as in *beet*) with spiky shapes (D'Onofrio, 2013; McCormick et al., 2015). In the largest-scale study to date, Westbury, Hollis, Sidhu, and Pexman (2018) examined the fit between 8,000 nonwords and roundness or spikiness. They found that the phonemes /oʊ/ (as in *boat*), /u/, /b/, /m/, and /ɑ/ (as in *bought*) were associated with roundness, and that the phonemes /t/, /k/, /z/, /i/, and /ɪ/ (as in *bit*) were associated with spikiness.[2]

[1] As discussed in Sidhu and Pexman (2018), we intend this term to refer to associations between phonemes and nonauditory properties. Thus, we exclude associations that arise from direct imitation. We also intend this term to refer to associations arising from properties of the phonemes themselves. Thus, we also exclude associations based simply on patterns in language (e.g., if /s/ were associated with plurality in English speakers). Our use of the term is consistent with what Hinton et al. (1994) termed *synaesthetic sound symbolism.*

[2] This refers to their phoneme-only model.

David Sidhu is now at the Department of Experimental Psychology at University College London.

✉ David M. Sidhu
dmsidhu@gmail.com

1 Department of Psychology, University of Calgary, Calgary, AB, Canada

2 Department of Psychology, University of Alberta, Edmonton, AB, Canada

3 Department of Computing Science, University of Alberta, Calgary, AB, Canada

Published online: 05 April 2021

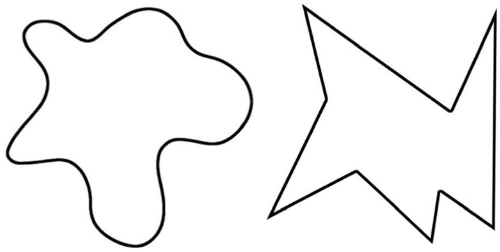

Fig. 1 Most people pair *maluma* with the left shape and *takete* with the right shape

Table 1 The terms *sound symbolism* and *iconicity*, as they will be used in this paper

Terms
Sound symbolic association: Associations between phonemes and perceptual and/or semantic properties. For example, *the maluma/takete effect*, in which certain sounds are associated with round or spiky shapes.
Iconicity: Resemblance between word form and meaning. This can include *onomatopoeic iconicity*, in which the sound of a word imitates some sound in the environment (e.g., *quack*). It can also include *cross-modal iconicity*, in which the sound of a word is associated with some nonauditory feature of the word's meaning—for instance, through sound symbolic associations (e.g., *balloon*).

Interest in sound symbolism is motivated in part by the long-debated relationship between linguistic form and meaning. The dominant view has been that this relationship is *arbitrary* (e.g., Hockett, 1963). One possible opposition to this is *iconicity.* In spoken language iconicity here refers to instances in which aspects of a word's form somehow resemble aspects of its meaning (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015).[3] The most obvious examples are when the sound of a word imitates a sound-based meaning (i.e., *onomatopoeia*; e.g., the sound of *quack* resembles duck sounds). However, as discussed, language sounds can also have associations with nonauditory properties (e.g., shape). This allows for the possibility of *cross-modal iconicity*. For instance, the phonemes in *balloon* are associated with roundness, allowing *balloon* to resemble its meaning via the associations of its component phonemes, making it iconic (see Table 1 for definitions).

While sound symbolic associations have been demonstrated using laboratory tasks with nonword stimuli, the presence of these associations in real language has not been established. That is, it is unknown whether there is a tendency for languages to contain cross-modally iconic words like *balloon.* Most investigations of cross-modal iconicity in real language have focused on a small number of words.[4] Many have investigated size iconicity. That is, whether the sound symbolic association between high-front vowels (e.g., /i/ as in beet) and small shapes, and low-back vowels (e.g., /ɑ/ as in *bought*) and large shapes (Newman, 1933; Sapir, 1929), is present in real langauge. One approach has been to analyze a sample of words in a language and see whether words for small or large things contain small-associated or large-associated phonemes. The results of this approach have been equivocal. Some studies find evidence of size iconicity in English (Thorndike, 1945). Some do not (Katz, 1986; Newman, 1933). Taking another approach, Blasi, Wichmann, Hammarström, Stadler, and Christiansen (2016) investigated the forms of 100 basic words across two-thirds of the world's languages. They found a tendency for the word meaning *small* to contain the phoneme /i/.

There has been less work exploring shape iconicity (i.e., the maluma/takete effect in language). Blasi et al. (2016) found that words meaning *round* tended to contain the phoneme /r/. They (and others; Johansson et al., 2020; Joo, 2020) have also found that across languages, words for some round body parts tend to contain round-associated phonemes (e.g., words for breast containing /m/, /u/). Katz (1986) explored the vowels contained in 325 English concrete words that had been rated on their shape (from round to angular). There was a trend in which words containing /u/ were the most round, though the overall effect of vowel type was not significant. More recently, Monaghan, Mattock, and Walker (2012) examined English words relating to roundness and angularity (311 and 198 words, respectively) Words for angularity were more likely to contain velars (e.g., /k/, /g/) and unvoiced consonants (e.g., /p/, /f/), but not after correcting for multiple comparisons.

Thus, whether shape iconicity is present in large samples of a lexicon is still unknown. This is an important question. Research on the maluma/takete effect in nonwords has proliferated in recent years, yet the relevance of this work to real language has remained an intriguing but unanswered question. In the present study we conducted the first large-scale investigation of shape iconicity in existing language. We first conducted an exploratory analysis to examine whether there are differences in the phonemes that tend to appear in English words for round versus spiky objects. Then, we conducted a confirmatory analysis to directly test whether round-associated phonemes (e.g., /oʊ/, /u/, /b/, /m/, and /ɑ/) are more common in English words referring to round objects, and spiky-associated phonemes (e.g., /t/, /k/, /z/, /i/, and /ɪ/; Westbury et al., 2018) are more common in words referring to spiky objects.

[3] While not the focus of the present work, we note that iconicity can also be present in spoken language beyond the word form. For example in prosody (e.g., rising pitch while saying "the bird was high in the sky") or co-speech gesture (e.g., holding the hands far apart while saying "the bird was huge").

[4] Two notable studies are recent large-scale investigations of phonemes in words of different valence (Adelman, Estes, & Cossu, 2018) and arousal (Aryani, Conrad, Schmidtke, & Jacobs, 2018).

Method

Participants

A total of 171 participants (43 males; M_{age} = 20.3 years, SD = 3.9; 84 at the University of Alberta and 87 at the University of Calgary) participated in exchange for partial course credit. All participants reported English fluency and normal or corrected-to-normal vision. We did not require that participants were native English speakers.

Materials and procedure

We chose to examine words referring to objects in order to have a large set of items (e.g., larger than if we had examined shape adjectives). We began with a list of more than 8,000 nouns in the CELEX database that had standardized concreteness ratings > 1.5 in Hollis, Westbury, and Lefsrud (2017). Hollis et al. derived these concreteness ratings by statistically extrapolating from 37,058 human-rated words (human ratings from Brysbaert, Warriner, & Kuperman, 2014, using a 5-point scale ranging from abstract to concrete) to a dictionary of 78,286 words, from a model that used Word2Vec vector values as predictors. We then conducted a pilot study in which a group of 67 participants (who did not participate in any other studies reported here) categorized subsets of these words as an "object" (e.g., *blanket*, *flag*, *mailbox*), "not an object" (e.g., *barber*, *herd*, *paisley*) or as an unknown word. We retained words that a majority classified as an object. We removed plurals and words referring to an object whose shape would be difficult to rate (e.g., mass nouns and objects without a defined shape, such as *veggie* or *gizmo*), to end up with 1,757 singular object nouns.

To obtain shape ratings of these objects we made use of best–worst ratings, which are more reliable and efficient than rating scales for collecting semantic norms (see Hollis, 2017; Hollis & Westbury, 2018; Kiritchenko & Mohammad, 2017). Participants were tested individually. Each rated 100 sets of six words. Participants saw six words at a time in the middle of a computer screen, presented using custom-written software. The order of the words, and the order of the 100 word sets, were random. On each trial, participants used the mouse to choose the words referring to the most round and the most spiky objects (see Fig. 2). Using software released by Hollis et al. (2017) for this purpose (available from https://sites.ualberta.ca/~hollis/), we ensured that the trial-to-word ratio was close to 8:1. Trials were constructed to contain minimal informational redundancy. Since every trial involved six words, this means that each word was expected to be judged 8 × 6 = 48 times. Previous work has found that a trial-to-word ratio of 8:1 is sufficient for reliable best–worst judgments (see Hollis, 2017; Hollis, 2019; Hollis & Westbury, 2018). Because we ended up with slightly fewer participants (171) than our goal (174), each word was judged on average 7.9 × 6 = 47.4 times. We used value scoring (explained in Hollis, 2017) to calculate spiky-round scores for each word based on these data (i.e., *shape rating*). The value-scoring algorithm predicts where an item should be rated relative to other items, based on its history of being chosen as rounder or spikier than other items, adjusting for the competitiveness of the items it appeared with (i.e., how often they were chosen as roundest or spikiest). Table 2 presents the 10 word referents judged as roundest and spikiest (all the ratings can be found at https://osf.io/nfkd2).

Results

Our initial exploratory analysis examined which phonemes were predictive of a word referent's shape. Our second analysis was confirmatory, using existing models of phonemes' associations with roundness and spikiness (Westbury et al., 2018) to quantify the sound symbolic associations of each word's sound, and to determine whether this predicted each word referent's shape.

Exploratory Analysis

The data were analyzed using least absolute shrinkage selection operator (LASSO) regression (see Friedman, Hastie, & Tibshirani, 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, Simon, Friedman, Hastie, & Tibshirani, 2011) which tunes the value of lambda for each of the predictors separately (see

Table 2 The words judged to have the 10 roundest-shaped and spikiest-shaped referents based on best–worst ratings

Ten roundest word referents	Ten spikiest word referents
Softball (3.42)	Spike (−3.34)
Ball (3.33)	Fork (−3.00)
Olive (3.15)	Porcupine (−2.81)
Pea (2.82)	Scalpel (−2.80)
Globe (2.74)	Swordfish (−2.64)
Tomato (2.73)	Tweezers (−2.56)
Balloon (2.73)	Switchblade (−2.53)
Grapefruit (2.72)	Pitchfork (−2.50)
Hoop (2.71)	Cactus (−2.49)
Donut (2.70)	Shrapnel (−2.43)

Note. *Z*-scored shape rating is shown in parentheses

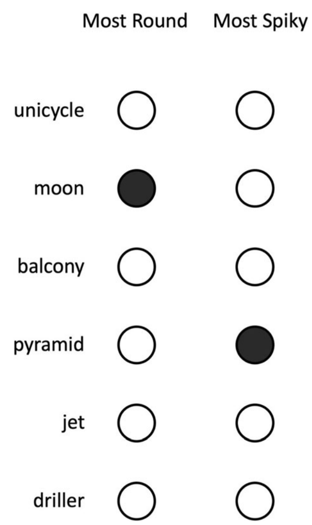

Fig. 2 An example trial. Participants were presented with six object words at a time and had to choose those that referred to the most round and most spiky objects

Zou, 2006). We used a common definition of lambda as the largest value of lambda that results in a model with mean cross-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. Annotated R code can be found at: https://osf.io/nfkd2.

The independent variables of interest were 39 dichotomous predictors that coded for the presence of every phoneme. We coded for the *presence* of each phoneme (rather than the *number* of each phoneme), because it was rare for any phoneme to appear more than once in a word. On average, each phoneme appeared more than once in only 0.75% of items[5] (see Fig. 3 for the frequency of each phoneme). We also included log-transformed word frequency +1 (Shaoul & Westbury, 2006), and number of phonemes (M = 5.07, SD = 1.53) as control variables. These variables were standardized and "forced" into the model by setting their lambda values to zero (i.e., they could not be removed by the LASSO process). The dependent variable was each word referent's shape rating. Thirty-nine words were excluded for not having available frequency values. One word was excluded because it was misspelled which left a total of 1,717 items.

The resulting model is summarized in Table 3. Nearly all of the phonemes that were more common in words referring to a round object in the present study have been shown in previous studies to have a sound symbolic association with roundness, except for /i/. Three of the phonemes that were more common in words referring to a spiky object (/k/, /t/, /ɪ/) have been shown in previous studies to have a sound symbolic association with spikiness. Additionally, the vowels /aɪ/ (as in *bite*) and /ɝ/ (as in *bird*) are consistent with the vowels typically associated with spiky shapes (i.e., front unrounded vowels). We also ran versions of this model only including noncompound nouns (n = 1,296), or monomorphemic words (n = 1,013; Balota et al., 2007). Three predictors (/i/[6], /ɪ/, /s/) no longer enter these models and thus should be interpreted with caution.

We ran analogous analyses predicting valence and arousal (n = 1,318; Warriner, Kuperman, & Brysbaert, 2013), concreteness (n = 1,620; Brysbaert et al., 2014; n = 1,717; Hollis et al., 2017), and size (n = 613; Scott, Keitel, Becirspahic, Yao, & Sereno, 2019) of our items. This was to determine whether phoneme presence would be predictive of *any* dimension, given a large set of items. No phoneme predictors entered any of these models.

Because results can differ by analysis method, we have included in Fig. 4 results from 12 other approaches to the main analysis (details on these analyses can be found in Electronic Supplementary Material here: https://osf.io/nfkd2). The general pattern is that most predictors of spikiness were robust to different analyses, while predictors of roundness were more variable. Results also differed depending on whether controls (in particular number of phonemes) were included.

Confirmatory analysis

We began by quantifying the sound symbolic association of each word's sound based on its component phonemes. This was based on models in Westbury et al. (2018), who asked participants to decide whether 8,000 nonwords were a good label for either an unspecified "round thing" or a "sharp thing" (i.e., on separate trials), and then generated two coefficients for each phoneme reflecting its association with roundness and spikiness. Using these coefficients, we computed the sound symbolic association between each of our words and both roundness and spikiness. This was done by summing the roundness coefficients for each phoneme in a word, and then doing the same for spikiness coefficients (see Fig. S1 for their distribution here: https://osf.io/nfkd2/). Sound symbolic roundness and spikiness were negatively correlated with one another ($r = -.52$, $p < .001$).[7] We summed these two values for each word (after reverse scoring spikiness scores) to create a single *sound symbolism score*. This quantifies the extent to which the *sound* of each word is associated with roundness or spikiness (see Table 4).

[5] We elected to only analyze phonology, at the exclusion of orthography, because of the high correlation between them. However, we have included analyses of orthography in Electronic Supplementary Material here: https://osf.io/nfkd2.

[6] We examined whether this was a result of /i/ appearing as a diminutive suffix but found that this suffix only appeared in eight of our words. Although these words tended to have a round shape ($M_{\text{Standardized Shape Rating}}$ = 0.91), we are hesitant to make any conclusions based on so few items. It is interesting to note that of the words containing an /i/ (n = 249), those that ended in an /i/ (not necessarily as a diminutive suffix) tended to have a rounder shape (n = 92; $M_{\text{Standardized Shape Rating}}$ = 0.47) than those that did not end in /i/ (n = 157; $M_{\text{Standardized Shape Rating}}$ = 0.01).

[7] Note that both sound symbolic roundness (b = 0.02, p < .001) and spikiness (b = −0.03, p < .001) are significant predictors of shape when either is used instead of the summed score in the subsequent analysis.

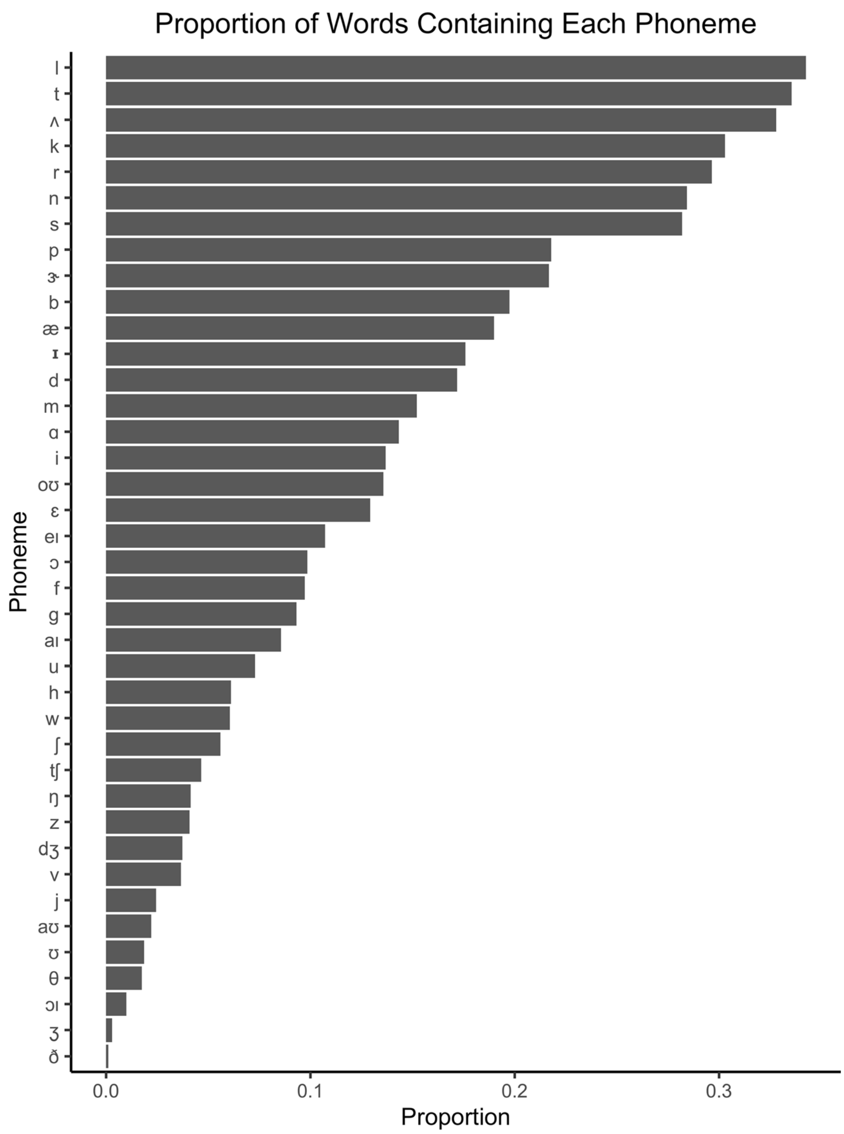

Fig. 3 The proportion of words used in the analysis that contained each of the phoneme predictors

We then ran a linear regression which included each word's sound symbolism score, number of phonemes, and logged frequency (Shaoul & Westbury, 2006) as predictors. All predictors were standardized. The dependent variable was each word referent's shape rating. This model revealed a significant effect of sound symbolism score ($b = 0.03$, $p < .001$; see Table 5). Words with round-associated sounds were more likely to refer to round objects, and words with spiky-associated sounds were more likely to refer to spiky objects. The patterns and significance were virtually unchanged if compound nouns or if multimorphemic words were excluded (in both cases $b_{\text{Sound}} = 0.03$, $p < .001$).

Finally, we examined the extent to which an alignment of sound symbolism score and shape rating (i.e., iconicity) agreed with existing subjective ratings of iconicity. We derived a measure of iconicity by multiplying standardized sound symbolism scores and shape ratings. Thus, items with both a round sound and shape (or spiky sound and shape) received positive values, while those showing a mismatch received negative values. There was a significant positive correlation ($r = .15$, $p < .001$) between our derived iconicity measure and existing subjective ratings of iconicity ($n = 577$; Perry, Perlman, & Lupyan, 2015; Winter, Perlman, Perry, & Lupyan, 2017).

Discussion

Associations between phonemes and shapes (i.e., the maluma/takete effect) have been well demonstrated with nonword stimuli in laboratory tasks. Here, we investigated whether these associations are attested in real language (i.e., whether round-associated or spiky-associated phonemes are more common in words referring to round or spiky objects, respectively), and thus whether shape iconicity is present in

Table 3 Predictors of word referent shape, separated into those predictive of object roundness (positive) and spikiness (negative)

Predictive of object roundness		Predictive of object spikiness	
Predictor	Coefficient	Predictor	Coefficient
u	0.021	aɪ	−0.036
m	0.018	tʃ	−0.035
oʊ	0.011	**k**	−0.029
b	0.010	ʃ	−0.026
i[a]	0.005	ɝ	−0.023
Frequency	0.003	r	−0.022
		Number of Phonemes	−0.016
		t	−0.011
		ɪ[a]	−0.010
		s[a]	−0.009

Note. Predictors whose coefficients shrunk to zero are not shown. Predictors whose coefficient signs are consistent with what has previously been found in laboratory tasks with nonword stimuli are shown in boldface. Coefficients reflect predicted change in object shape (which ranges from 0 = *extremely sharp*; 1 = *extremely round*) from the presence of each phoneme

[a] Predictor does not enter model when excluding compound nouns or multimorphemic words

Table 4 The 10 words from our stimuli whose sounds are most sound symbolically associated with roundness and spikiness, based on the phoneme-only model in Westbury et al. (2018)

Ten roundest-sounding words	Ten spikiest-sounding words
Sombrero (3.64)	Cheesecake (−2.45)
Doberman (3.17)	Knickers (−2.36)
Bamboo (3.14)	Cleat (−2.17)
Broom (3.14)	Cutlery (−2.17)
Trombone (3.09)	Factory (−2.17)
Armband (2.77)	Parakeet (−2.17)
Blossom (2.77)	Taxi (−2.17)
Bomber (2.77)	Turkey (−2.17)
Marble (2.77)	Worksheet (−2.17)
Snowmobile (2.63)	Anklet (−2.08)

Note. Z-scored sound symbolic association score in parentheses

language. The exploratory analysis showed that four phonemes with sound symbolic associations to roundness (/u/, /m/, /oʊ/, /b/) were more common in words referring to round objects, and three phonemes with sound symbolic associations

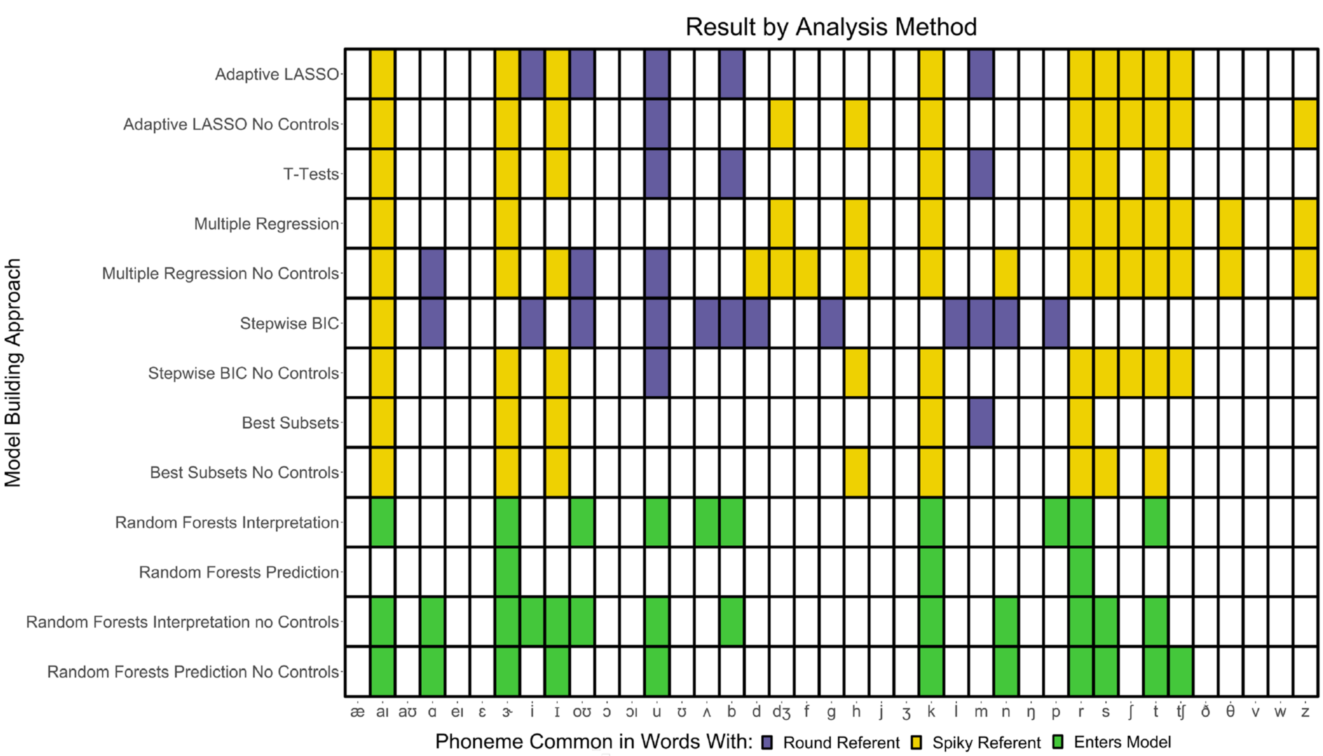

Fig. 4 Results of 13 different approaches to the main exploratory analysis. Each analysis examined whether the presence of a given phoneme predicted a word's shape rating. Cells in purple convey that a given phoneme was found to be predictive of a word referring to a round object; those in yellow were predictive of a word referring to a spiky object. These different methods included Adaptive LASSO (as reported in the main analysis), Bonferroni-corrected *t* tests, multiple regression, stepwise regression including forward and backward changes using Bayesian Information Criterion (BIC), best subsets regression which selected the model of all possible models resulting in the lowest BIC, and two stages of a random forest analysis which calculated the importance of variables across many randomly constructed models (note that the output of these models only indicate whether predictors warrant inclusion in the model, not their coefficients). Full details can be found in Electronic Supplementary Material here: https://osf.io/nfkd2

Table 5 Linear model predicting each word's shape

Predictor	*b*	*SE*	*t*	sr^2	*p*
Intercept	0.50	0.003	185.93		<.001***
Sound symbolism score	0.03	0.003	10.41	0.06	<.001***
Number of phonemes	–0.02	0.003	–7.14	0.03	<.001***
Frequency	0.005	0.003	1.55	0.00	.12

Note. Adjusted R^2 = 0.10. ***$p < .001$.

to spikiness (/k/, /t/, /ɪ/) were more common in words referring to spiky objects. The phonemes /aɪ/ and /ɝ/ were also more common in words referring to spiky objects, which is consistent with front unrounded vowels' sound symbolic associations with spikiness. Further, when we directly quantified the sound symbolic association of each word's phonology (as round-associated or spiky-associated), this predicted the shape of a word's referent.

Two of the vowel phonemes more common in words referring to round objects (/u/ and /oʊ/) were back and rounded vowels. Front and unrounded vowels (/aɪ/, /ɝ/, and /ɪ/) were more common in words referring to spiky objects. The reverse of this pattern was observed in the phoneme /i/ being more common in words for round objects. Note that this association was not present when compound nouns or multimorphemic items were removed. The consonants that were more common in round objects (/m/ and /b/) were both voiced bilabials. Contrasting them with the consonants more common in spiky objects (e.g., /t/ and /k/) reveals a smoother sound and articulation for /m/ and /b/, both of which could be associated with roundness (and the converse with spikiness) through perceptuomotor analogy.

Unexpectedly, the affricate /tʃ/, fricatives /ʃ/ and /s/, and the approximant /r/,[8] were all more common in words for spiky objects. Winter (2016) found that the phoneme /r/ was more common in words denoting rough versus smooth textures. Aryani et al. (2018) found that words containing hissing sibilants (e.g., /s/ and /ʃ/) tended to be higher in affective arousal. Both the rough/smooth and excited/calm dimensions could partially overlap with the spiky/round dimension, and thus help explain these results. Notably, these phonemes, along with the other consonants common in words for spiky objects, all involve the tongue in their articulation, while those common in words for round objects are all articulated with the lips. We note the somewhat contradictory finding by Blasi et al. (2016) that /r/ is common in words meaning "round" across languages.

[8] While we use /r/ for convenience here, we do not intend to suggest a trill, but rather the approximant /ɹ/, as is typical in most American dialects.

Words referring to spiky objects tended to contain more phonemes. This persisted even after accounting for age of acquisition (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) in a supplementary analysis. Part of the explanation could be that spikier objects in our dataset also tended to be larger (n = 613; r = −.26, $p < .001$), since words for larger objects tended to contain more phonemes (r = .13, $p < .001$). Thus, the relationship between length and spikiness could be indicative of an iconic relationship between word length and referent size. A mediational analysis found evidence that size partially mediated the relationship between length and shape (Average Causal Mediation Effect = −0.006, $p < .001$; Average Direct Effect = −0.04, $p < .001$). Previous work has also demonstrated that participants will associate longer nonwords with more visually complex objects (Lewis & Frank, 2016). Thus, the relationship between length and spikiness could reflect a greater visual complexity in spikier objects.

Our main finding was that many of the associations between phonemes and shapes found in laboratory tasks are attested in the pairing between sound and meaning in English. Certainly, this is a modest effect—phoneme predictors and sound rating had small coefficients in the exploratory and confirmatory analyses, respectively. Many other factors play larger roles in the form of language. Nevertheless, the presence of shape iconicity was observable in the present analyses. The process by which this occurs is unknown. From an evolutionary standpoint, aspects of language that convey some benefit to processing and learnability should have an advantage and thus survive (Monaghan, Christiansen, & Fitneva, 2011). Iconic forms can be easier to learn and remember (e.g., Imai, Kita, Nagumo, & Okada, 2008; Lockwood, Dingemanse, & Hagoort, 2016), perhaps affording them a slight lexical advantage. Future historical studies will be necessary to understand this process.

The opposite interpretation is that shape iconicity in a language creates the maluma/takete effect observed in laboratory tasks with nonwords (see Taylor, 1963). While it is not possible to rule this out, we believe it is unlikely that the patterns observed here are *entirely* responsible for effects with nonwords. This is because the maluma/takete effect has been observed in speakers of different languages (see Styles & Gawne, 2017). If patterns in language created those effects, then shape iconicity would have had to emerge accidentally in each of the languages in which the maluma/takete effect has been observed. However, it is possible that patterns in language may *strengthen* sound symbolic associations, creating a feedback loop (see *language specific iconicity* in Imai & Kita, 2014). An interesting topic for future research would be to examine whether the extent to which shape iconicity is observed in a language predicts the strength of the maluma/takete effect in its speakers.

Another possibility is that words' sounds affected best–worst ratings. However, the predictive effect of sound score (i.e., the confirmatory analysis) was diminished for items in the middle two quantiles of shape ratings (i.e., those with more ambiguous shapes; $b = 0.00$, $p = .048$) compared with words at the extremes ($b = 0.05$, $p < .001$). This is not consistent with sound symbolism driving shape ratings, as these ambiguous shapes should be more susceptible to the effect of sound.

The present results are also indicative of another type of nonarbitrariness, namely *systematicity*: large-scale patterns in the forms of words belonging to the same syntactic or semantic category (Dingemanse et al., 2015); in this case objects of a certain shape. While systematic patterns need not be iconic, the two are not mutually exclusive. The present pattern is perhaps best described as *systematic iconicity*: there are patterns in the forms of words belonging to the categories of round and spiky, and the specific nature of those patterns is iconic. Since systematicity tends to be pervasive in a language (Dingemanse et al., 2015), systematic iconicity could represent a means by which iconicity is broadly relevant to a lexicon, beyond specific classes of words (e.g., onomatopoeia).

The maluma/takete effect has been studied for nearly 100 years. In that time, its connection with real language has been largely unexplored. The present results suggest that the maluma/takete effect is attested in the English lexicon, in the form of shape iconicity.

Acknowledgements This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery Grants to C.W. and P.M.P. The authors would like to thank Kristen Deschamps for her help collecting data.

Contributions All authors participated in the design and interpretation of reported results, the analysis of data, and in drafting and revising the manuscript.

Open practices statement All data and analysis codes are available at: https://osf.io/nfkd2/

The experiment was not preregistered.

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

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, *63*, 196–205. https://doi.org/10.1016/j.cortex.2014.08.025

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/bf03193014

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America, 113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

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

D'Onofrio, A. (2013). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech*, *57*, 367–393. https://doi.org/10.1177/0023830913507694

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

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, *21*, Article e12659. https://doi.org/10.1111/desc.12659

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

Hinton, L., Nichols, J., & Ohala, J. J. (1994). Sound-symbolic processes. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 1–14). New York, NY: Cambridge University Press.

Hockett, C. (1963). The problem of universals in language. In J. Greenberg (Ed.), *Universals of language* (pp. 1–22). Cambridge, MA, MIT Press.

Hollis, G. (2017). Scoring best–worst data in unbalanced many-item designs, with applications to crowdsourcing semantic judgments. *Behavior Research Methods, 50*(2), 711–729. https://doi.org/10.3758/s13428-017-0898-2

Hollis, G. (2019). The role of number of items per trial in best–worst scaling experiments. *Behavior Research Methods,* 1–29. https://doi.org/10.3758/s13428-019-01270-w. Advance online publication.

Hollis, G., & Westbury, C. (2018). When is best–worst best? A comparison of best–worst scaling, numeric estimation, and rating scales for collection of semantic norms. *Behavior Research Methods*, *50*(1), 115–133. https://doi.org/10.3758/s13428-017-1009-0

Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *The Quarterly Journal of Experimental Psychology*, *70*(8), 1603–1619. https://doi.org/10.1080/17470218.2016.1195417

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 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*, 54–65. https://doi.org/10.1016/j.cognition.2008.07.015.

Johansson, N. E., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology, 24*, 253–310.

Joo, I. (2020). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. *Linguistic Typology, 24*, 1–12.

Katz, A. N. (1986). Meaning conveyed by vowels: Some reanalyses of word norm data. *Bulletin of the Psychonomic Society, 24*, 15–17. https://doi.org/10.3758/BF03330490

Kiritchenko, S., & Mohammad, S. M. (2017). Best–worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. *Proceedings of The Annual Meeting of the Association for Computational Linguistics* (pp. 465–470). http://www.aclweb.org/anthology/P/P17/

Köhler, W. (1929). *Gestalt psychology*. New York, Liveright.

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

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195. https://doi.org/10.1016/j.cognition.2016.04.003

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1274–1281. https://doi.org/10.1037/xlm0000235

McCormick, K., Kim, J. Y., List, S., & Nygaard, L. C. (2015). Sound to meaning mappings in the bouba–kiki effect. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1565–1570). Austin, TX: Cognitive Science Society.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, *140*(3), 325–347. https://doi.org/10.1037/a0022924

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152–1164. https://doi.org/10.1037/a0027747

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology*, *45*, 53–75. https://doi.org/10.2307/1414186

Nielsen, A. K. S., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology*, *65*, 115–124. https://doi.org/10.1037/a0022268

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

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

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. (2006). USENET orthographic frequencies for 1,618,598 types. (2005–2006) Edmonton, AB: University of Alberta. http://www.psych.ualberta.ca/~westburylab/downloads/wlallfreq.download.html

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

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software, 39*(5), 1–13. http://www.jstatsoft.org/v39/i05/

Styles, S. J., & Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter *i-Perception*, *8,* 1–17. https://doi.org/10.1177/2041669517724807

Taylor, I. K. (1963). Phonetic symbolism revisited. *Psychological Bulletin, 60*, 200–209. https://doi.org/10.1037/h0040632

Thorndike, E. L. (1945). On Orr's hypotheses concerning the front and back vowels. *British Journal of Psychology, 36*, 10–14.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 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. (2016). *The sensory structure of the English lexicon* (Doctoral dissertation). https://escholarship.org/uc/item/885849k9

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

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