COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) e12791

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12791

What Does a Horgous Look Like? Nonsense Words Elicit Meaningful Drawings

Charles P. Davis,[a,b,c] Hannah M. Morrow,[a,b] Gary Lupyan[d]

[a]Department of Psychological Sciences, University of Connecticut
[b]Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut
[c]Brain Imaging Research Center, University of Connecticut
[d]Department of Psychology, University of Wisconsin-Madison

Received 1 April 2019; received in revised form 29 August 2019; accepted 3 September 2019

Abstract

To what extent do people attribute meanings to "nonsense" words? How general is such attribution of meaning? We used a set of words lacking conventional meanings to elicit drawings of made-up creatures. Separate groups of participants rated the nonsense words and the drawings on several semantic dimensions and selected what name best corresponded to each creature. Despite lacking conventional meanings, "nonsense" words elicited a high level of consistency in the produced drawings. Meaning attributions made to nonsense words corresponded with meaning attributions made by separate people to drawings that were inspired by the name. Naïve participants were able to recover the name that inspired the drawing with greater-than-chance accuracy. These results suggest that people make liberal and consistent use of non-arbitrary relationships between forms and meanings.

Keywords: Iconicity; Sound symbolism; Semantic; Abstraction

1. Introduction

Imagine a group of artists illustrating children's books about made-up creatures with names like "horgous" and "keex." Will different artists create similar drawings for creatures that have similar names? Will readers who encounter the drawing of a "horgous" expect this creature to be named "horgous"? We show here that drawings elicited by certain "nonsense" words relate in a systematic way to the form of these words. This relationship between nonce words and the meanings they express is bidirectional: Certain wordforms lead people to infuse their drawings with certain properties. Other people,

Correspondence should be sent to Charles P. Davis, Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Storrs, CT 06269. E-mail: charles.davis@uconn.edu

looking at the drawings, match them back to the original wordforms at higher-than-chance levels.

The idea that certain words fit some meanings better than others has its roots in the ancient world (Plato, 1999), but it was all but excised by structural linguistics (de Saussure, 1959) and its focus on the sharp boundary between the signifier and the signified. Conventional wisdom has been that with the exception of words that directly imitate sounds, the relationship between word-forms and meanings is arbitrary: "There is no reason for you to call a dog 'dog' rather than 'cat' except for the fact that everyone else is doing it" (Pinker & Bloom, 1990, p. 728).

In an early systematic investigation of what he called "phonetic symbolism," Sapir (1929) presented people with dozens of short nonce words and asked them to distinguish the words on size. For example, a participant may be told that "mal" and "mil" both mean table; they then had to decide which would be a better word for a *large* table. The chosen answer, overwhelmingly and largely independent of age and language background, was "mal," such that 75%–96% of people prefer "mal" to describe a *large* table.[1] Sapir speculated that these sound-to-meaning mappings may arise from people implicitly learning that producing certain vowels requires larger mouth cavities. This early speculation was amplified by Ramachandran and Hubbard's (2001) replication of Sapir's phonetic symbolism demonstration (see also Köhler, 1929; Newman, 1931), giving us the well-known "bouba-kiki" effect wherein people overwhelmingly match "bouba" to a round shape and "kiki" to an angular one (see also, e.g., Maurer, Pathman, & Mondloch, 2006). Further, both vowels and consonants seem to contribute to the tendency to match certain sounds (e.g., /m/, /u/) with certain shapes (e.g., rounded figures; Ahlner & Zlatev, 2010; see also D'Onofrio, 2014; Nielsen & Rendall, 2011, 2013), and the effect exists independently in both spoken and written language (i.e., via the curvature of round-sounding letters like /o/; Cuskley, Simner, & Kirby, 2017). There is also a graded relationship between sound and size: Increasingly large-sounding nonce words are associated with increasingly large objects (Thompson & Estes, 2011).

In the last several decades, iconicity—a resemblance between form and meaning—has been increasingly recognized as a basic design feature of natural language in both the signed and spoken modalities (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Monaghan, Shillcock, Christiansen, & Kirby, 2014; cf. Hockett, 1978; Perniss & Vigliocco, 2014). The idea that the auditory modality can convey meanings in an iconic way—beyond simple imitation of sounds—is at first counterintuitive. For example, Hockett argued that the relationship between spoken words and meanings is arbitrary because "When a representation of some four-dimensional hunk of life has to be compressed into the single dimension of speech, most iconicity is necessarily squeezed out" (Hockett, 1978, p. 274). We now recognize that speech is a richly multi-dimensional signal, and spoken languages make ample use of this dimensionality to convey meanings in an iconic way. For example, languages make systematic use of consonant voicing (/b/ vs. /p/, /d/ vs. /t/) to signal differences in mass: Siwu: tsratsra, "a light person walking quickly" vs. dzradzra, "a heavy person walking quickly," where voiceless consonants like /t/ correspond to lightness and voiced consonants like /d/ correspond to heaviness. Vowel quality is used to signal size: Ewe: lɛgɛɛ: logoo, "slim: fat." Vowel lengthening is used to signal duration and intensity: Japanese: piQ: piiQ, "tear short: long strip of cloth." Reduplication is used to signal repetition: Tamil: curuk-nu: curukcuruk-nu, "a sharp prick: many sharp pricks" (Dingemanse et al., 2015; see also Perniss, Thompson, & Vigliocco, 2010). In some languages, vowel height or frontness is used systematically to mark diminutives (Ultan, 1978).

Although these form-to-meaning relationships (what Dingemanse et al., 2015, call "relative iconicity") are not found in all languages, examining statistical relationships between forms and meanings across languages does reveal some more universal relationships (what Dingemanse et al., 2015, refer to as "absolute iconicity") such as the higher likelihood of using sounds /i/, /C/ in words for "small" and the sound /r/ in words for "round" (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; see also Dautriche, Mahowald, Gibson, & Piantadosi, 2017). Both relative and absolute iconicity are legitimate forms of iconicity because for both, a speaker can infer something about the meaning of a word from aspects of its form.

What makes the examples of relative iconicity described above especially interesting is that people appear to be sensitive to such form–meaning relationships even when they are not phonemically expressed in their language. For example, in English, smaller objects do not, as a rule, have shorter names than larger objects,[2] yet when asked to select a nonce-word for a small object such as a pin, people not only prefer shorter words, but justify their choices with statements like "a small item's name should be small" and "pins are sharp and simple, as is this word" (Lupyan & Casasanto, 2012). While Japanese has over 1,700 "sound-symbolic" words, many of them in common use (Allen et al., 2007), English does not. Yet monolingual English 3-year-olds are sensitive to some of Japanese form–meaning relationships when learning novel words (Kantartzis, Imai, & Kita, 2011). And although English words for small animals do not, as a rule, employ higher-pitched sounds, there are robust cross-modal links between size and auditory pitch: larger animals tend to be associated with low-pitched sounds and vice versa (Ohala, 1994). This may help explain why, when reading a children's book aloud, there is something natural about saying "elephant" in a lower pitch than "mouse," or using a higher pitch to refer to the baby elephant compared to the mommy elephant.[3] Such vocal (and gestural) iconicity is on full display in the popular *Baby Shark* song, where baby/mommy/daddy shark is depicted by progressively lower pitch and larger arm and hand movements. Grandma and grandpa shark age is depicted by an inward rounding of the fingers and simultaneous rounding of the vowels to depict dentures.[4]

Hearing adults with no sign language experience can also make inferences about the meaning of sign language gestures to determine some quite subtle aspects of meaning such as distinguishing whether a gesture refers to an event with a finite end point or not. For example, presented with a sign for "think," people are more likely to choose "believe" (similar telic content) over "forget" (Strickland et al., 2015).

Although these investigations of sound symbolism have not settled the question of where these associations between forms and meanings come from (but see Imai & Kita, 2014; Sidhu & Pexman, 2018; Spence, 2011), they have further demonstrated the varied way in which iconicity plays a role in language learning and vocal communication. For example, Perry, Perlman, Winter, Massaro, & Lupyan, (2018) showed that more iconic words are learned earlier by children (adjusting for numerous potential confounds like frequency, concreteness, and communicative need; see also, e.g., Imai, Kita, Nagumo, & Okada, 2008; Maurer et al., 2006; Peña, Mehler, & Nespor, 2011; for further review, see Imai & Kita, 2014). Such apparent advantages of iconicity go beyond word-learning. For example, people think that Bob is a better name than Mike for a round figure at rates well above chance level (Sidhu & Pexman, 2015). Further, Lupyan and Casasanto (2015) had people learn to categorize two kinds of "aliens." The aliens in one of the categories were subtly more round and in the other more pointy. When the categories were labeled with the nonce words "foove" (which people tend to associate with being round and friendly) and "crelch" (pointy and dangerous), people learned the category distinction itself (not just the category names) better than when arbitrary or iconically incongruent labels were used. When tasked with creating novel vocalizations to communicate a range of meanings (e.g., big, small, high, low, smooth, rough, cook, fire, fruit, and many others), people not only converge on surprisingly similar vocal forms, but when these vocalizations are played to naïve listeners (including those from other language backgrounds), they are understood at levels well above chance (Perlman & Lupyan, 2018; see also Parise & Pavani, 2011; Perlman, Dale, & Lupyan, 2015).

1.1. The present study

Prior work has provided ample evidence that certain seemingly nonsense words are nevertheless imbued with meaning: asked for the meaning of such a word, people make similar choices in forced-choice tasks (e.g., Maurer et al., 2006; Ramachandran & Hubbard, 2001; Sidhu & Pexman, 2015; Thompson & Estes, 2011), and asked to produce a novel vocalization to communicate a meaning, people create non-arbitrary vocalizations (e.g., Perlman et al., 2015), further indicating that people have expectations about the "natural" relationship between certain forms and meanings. Our main goal here is to investigate the generality of these non-random associations by using an open-ended task —free drawing. Returning to the thought experiment from our opening paragraph: If people are asked to draw a creature named with a nonce word, are aspects of meaning elicited by the nonce word infused into the drawing in a way that can be recovered by people viewing the drawings? Can people match the drawings back to the words that initially inspired them?

The full task sequence is schematized in Fig. 1. We first asked people to rate the extent to which a set of nonce words connote several properties (*infer a meaning task*; Fig. 1A). After collecting data for the *infer a meaning task*, we pre-registered two predictions and the methods to test them (https://osf.io/7wfxj/). First, the property ratings derived from the *infer a meaning* task should correlate with property ratings on the drawing task (Fig. 1B). For example, nonce words depicting roundness should elicit more round creatures. This should also be the case for gender, a more abstract property (see Westbury, Hollis, Sidhu, & Pexman, 2018). Nonce words that depict femininity should elicit more feminine-looking creatures. Second, the properties from the *infer a meaning* task (e.g., roundness) should predict not only the roundness of the drawn creature, but may also generalize to more abstract properties. Specifically, we predicted that "spikiness" (or *sharp*ness) would correlate with a rating of the creature's "intelligence" and the rating of largeness with a rating of dominance (see Auracher, 2017). This would suggest that non-arbitrary mappings between form and meaning are used not only to infer the concrete properties of novel creatures, but also more abstract properties like social dominance. After submitting this pre-registration, but before collecting the data for tasks C and D (Fig. 1), we also sought to determine whether people would be able to recover the word used to elicit the original drawing from the drawing itself and whether people's choices of picture-name were predictable from the iconic properties of the choices.

2. Materials and methods

2.1. Participants

Participants completed at most one of the tasks shown in Fig. 1. In the *infer a meaning* task (Fig. 1A), people were asked to infer meanings of nonwords (N = 151). In the *draw a creature* task (Fig. 1B), people drew creatures in response to these nonwords (N = 22). In the *infer properties* task (Fig. 1C), people inferred properties of the creature drawn in task B (N = 230). In the *recover the creature name* task (Fig. 1D), people named the drawn creature (N = 210). All participants were US-based native English speakers recruited from Amazon Mechanical Turk. They were compensated 0.45 USD for their participation. The procedure was approved by the University of Wisconsin-Madison institutional review board.

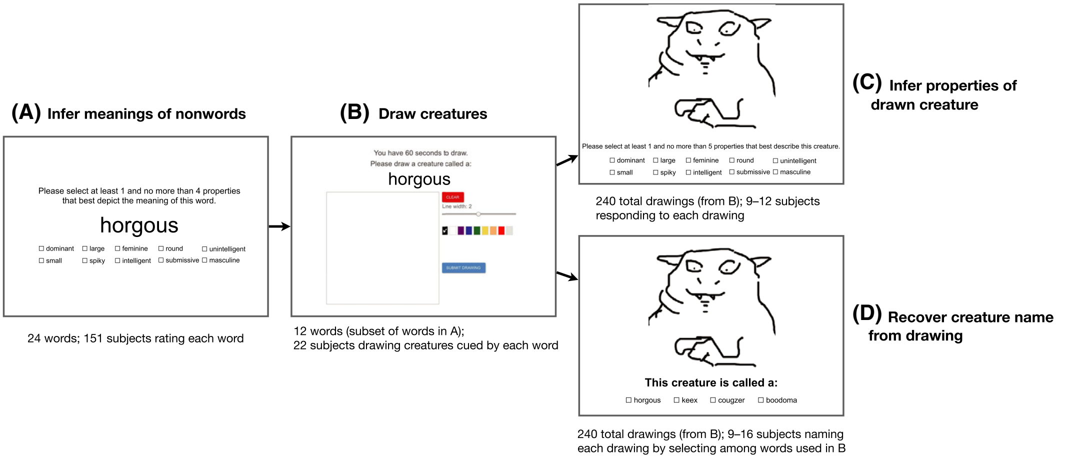

Fig. 1. Experimental protocol. Four separate groups of participants completed the four tasks shown. (A) Infer a meaning: Given a nonce word, select properties that it connotes. (B) Draw a creature: Given a subset of words from A, draw a creature that would be named by that word. (C) Infer properties: Given a creature drawn in B, rate its properties. (D) Recover the creature name: Given a creature drawn in B, select its name from among those used in B.

2.2. Power analysis

We conducted a power analysis using G*Power based on previously collected data in which people rated drawings of knives drawn to one of two prompts: “Draw a knife called a *teetay*” or “Draw a knife called a *tukeetee*” (people drawing “tukeetee” knives were expected to draw sharper knives than people asked to draw “teetay” knives). In that study, there were 26 total drawings and 12 raters per drawing; each rater saw both “teetay” and “tukeetee” knives (Lupyan & Casasanto, 2012). Teetay knives were rated as being less sharp, a mean difference of 0.7 (on a 7-point Likert scale), $SD_{\text{difference}} = 0.46$. We anticipate a smaller effect here (assuming $d = 0.7$), but power is increased in this study (by an uncertain amount) by having more trials per subject. Power analysis is further complicated by the presence of both random subject and item effects. With an effect size of $d = 0.7$ and power of 0.9, we require 21 subjects (16 for 0.8 power). We recruited enough participants to obtain 20 drawings per word and 16 ratings per drawing, per task.

2.3. Procedure

2.3.1. Infer a meaning

We began by extracting 24 nonce words from Westbury et al. (2018), who examined how various phonological features, phonemes, and letters were associated with various semantic dimensions such as roundness, size, and gender. We selected our initial set of words based on their likelihood of representing roundness, spikiness, largeness, smallness, masculinity, and femininity based on Westbury et al.’s (2018) published norms. Because our main goal was to determine the effects that nonce words can have on people’s behavior, we chose to use as materials nonce words that have been previously demonstrated to convey certain meanings. We presented this initial set of words to a group of naïve participants ($N = 157$) who were asked to rate each word on a set of eight properties (rounded, spiky, large, small, masculine, feminine, hot, and cold), choosing between 1 and 4 of the properties they thought best described what the word might mean. The task included several catch trials on which participants were instructed to select two specific options (e.g., masculine and cold). Six participants failed to correctly respond to the catch trials and were excluded from further analysis, leaving $N = 151$ (Fig. 1A).

Based on the results of our initial norming study, we selected 12 words that varied most systematically on the six target properties used in the main experiment. To select these 12 words, we computed the frequency with which participants selected each property. *Hot* and *cold* were not included in this selection process because they were only included in the word rating task as distracter options. As detailed in our pre-registration document, we had no *a priori* predictions about the relationship between the selected words and these two properties (to our knowledge, there is no prior work investigating temperature-related sound symbolism). These property counts were then z-transformed within each property (e.g., for the proportion of “large” responses, proportion of “small” response, and so on) to create a standard scale. We visually inspected the words in a space defined by three dimensions ($z_{\text{round}} - z_{\text{spiky}}$; $z_{\text{large}} - z_{\text{small}}$; $z_{\text{masculine}} - z_{\text{feminine}}$)

and selected 12 words that occupied different regions of the space as experimental stimuli. This selection method can also be thought of as choosing words that minimized the correlations among the three property dimensions (though as shown in Table 1, some of the correlations remained high). The selected words were *ackie*, *ambous*, *axittic*, *bomburg*, *boodoma*, *cougzer*, *cruckwic*, *flissil*, *gricker*, *heonia*, *horgous*, and *keex* (see Fig. 2). The correlations among the word rating properties are shown in Table 1.

2.3.2. *Draw a creature*

Each of the 12 words was presented in random order to participants in a drawing task ($N = 22$; Fig. 1B). Participants in the drawing group were given the following instructions:

> You will be asked to draw some creatures. Some of them correspond to real animals (like a rabbit). Most, however, are imaginary. For example, you might be asked to draw a "sask." Sound out the word and use your imagination about what a creature with this name would look like. You will have 60 seconds from the time you begin drawing to complete each drawing.

The drawing applet allowed participants to draw on a white background with black, red, blue, green, yellow, and purple pens, and vary the line thickness. This allowed participants maximal flexibility to represent the properties of interest (large–small, round–spiky, masculine–feminine). The created drawings made up the stimuli for the two experimental tasks described below.

Participants were also asked to draw a picture of a dog and a cat as attention checks. The drawings of those who failed the to draw both a dog and a cat, as determined through visual inspection of the drawings for four-legged creatures with some resemblance to a dog and cat, regardless of artistic quality, were not included in subsequent studies ($n = 2$).

2.3.3. *Infer properties of drawn creature*

The drawings were shown to a separate group of participants ($N = 210$; Fig. 1C) who saw the 12 drawings from one randomly chosen drawer to reduce effects of artistic quality within participants. The drawings were shown one at a time in random order. For each

Table 1
Pearson's correlations between property ratings

	12 Stimulus Words			24 Original Words		
	Round–Spiky	Large–Small	Masculine–Feminine	Round–Spiky	Large–Small	Masculine–Feminine
Round–spiky	1.00	—	—	1.00	—	—
Large–small	0.73	1.00	—	0.68	1.00	—
Masculine–feminine	0.21	0.46	1.00	0.45	0.69	1.00

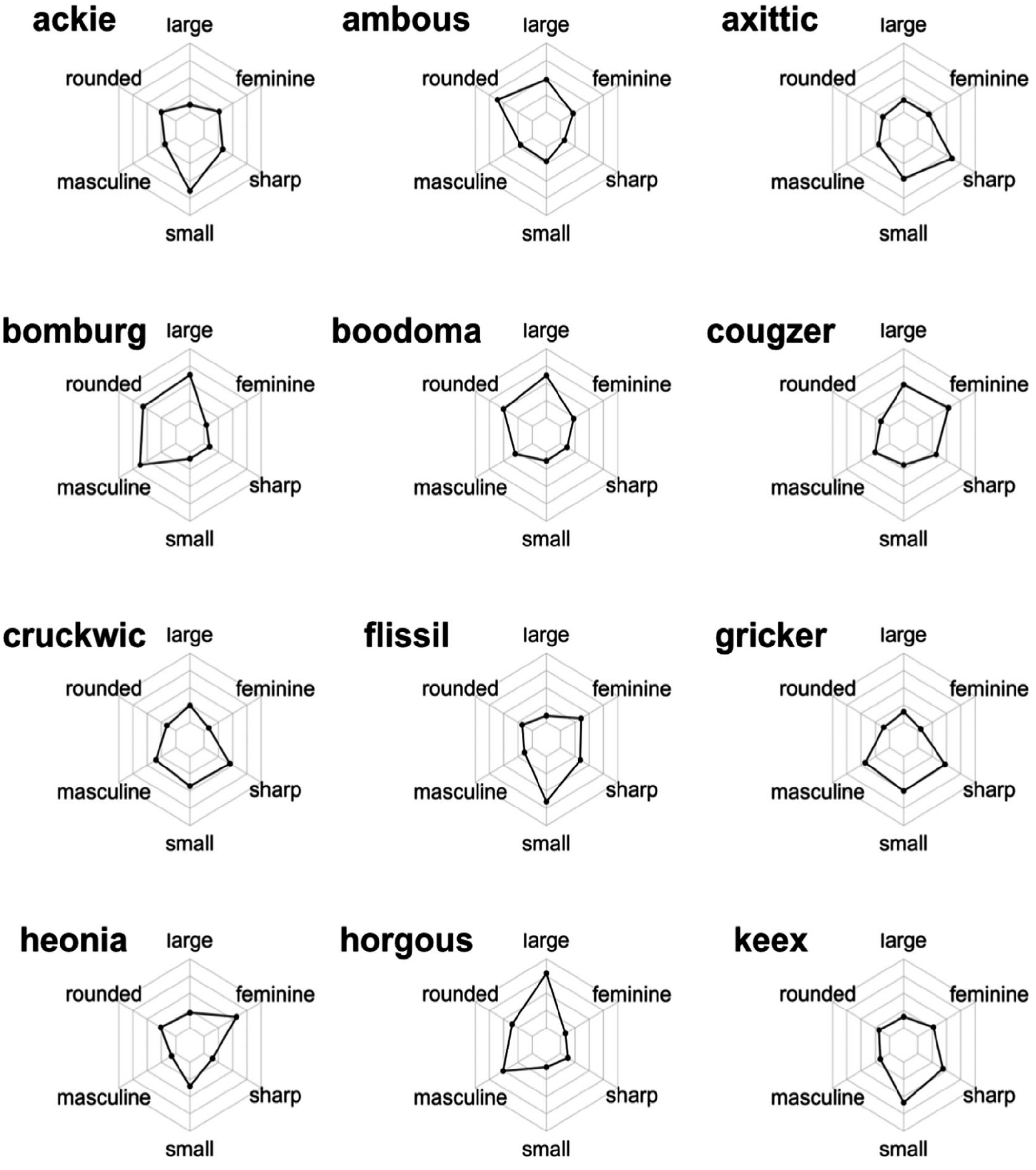

Fig. 2. Radar charts showing the characteristics (round–spiky, large–small, masculine–feminine) of each word selected as stimuli. The properties, shown as points in the radar space ranging from 0 to 1, are represented as proportions (*n* times property was selected/*n* total rating observations for that word).

drawing, participants were shown 10 properties (rounded, spiky, large, small, masculine, feminine, intelligent, unintelligent, dominant, submissive) and asked to select which properties best characterized the drawing by selecting between 1 and 5 properties from the list.

2.3.4. Recover creature name from drawing

Participants ($N = 230$) saw the drawings each presented with four of the non-words (Fig. 1D). The four words were the correct target word actually used to elicit the drawing (e.g., *horgous*), a highly dissimilar word (i.e., embodying the opposite sound-symbolic properties, where if the target was rated as highly large and round, the highly dissimilar word was rated as highly small and spiky; e.g., *keex*), a similar word (i.e., with similar sound-symbolic properties, e.g., *bomburg*), and an unrelated word (e.g., *cougzer*).[5] Each participant again was presented with drawings from a single drawer. Thus, participants each saw 12 drawings, with each drawing presented with a different set of four words. They were instructed to choose the word that best fit the drawing they saw.

2.4. Data analysis

Data were analyzed using logistic mixed effects regression models in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R. The model for the naming task was constructed as follows, with a treatment-coded fixed effect for word type (with *correct* as the reference class, and *similar*, *unrelated*, and *opposite* as the other options) and random effects for person who made the drawing, the person rating the drawing, and the word used in the prompt. The dependent variable (*is_type_chosen*) simply reflects whether, on a given trial, a given word type was chosen (thus, this could either be 1 or 0).

$$is\_type\_chosen \sim word\_type + (1|drawer) + (1|rater) + (1|word)$$

The models for the rating task, which are outlined in an OSF preregistration document (https://osf.io/7wfxj/), were constructed in the following form:

$$\begin{aligned} &is\_property\_chosen \sim (round\_rating - sharp\_rating) \\ &\quad + (large\_rating - small\_rating) + (masculine\_rating - feminine\_rating) \\ &\quad + (1|drawer) + (1|rater) + (1|word) \end{aligned}$$

We constructed six models to test whether the drawings incorporated the iconic properties of the words that elicited them (one model for each of spiky, round, large, small, masculine, and feminine), where the dependent measure (*is_property_chosen*) is a 1 or 0 reflecting whether, on a given trial, that property was chosen and the predictor variables are the z-scored property measures described in Section 2.3.1. We then tested whether the drawings carried *abstract* properties as predicted by the iconic properties of the words (i.e., spikiness/roundness is associated with intelligence while largeness/smallness is associated with dominance). This was done using four additional models, testing for the likelihood of choosing intelligent, unintelligent, dominant, and submissive as properties.

3. Results

3.1. Naming task

Did participants actually name the creature using the word that elicited its drawing? Yes—as shown in Fig. 3A, participants viewing the drawings were more likely to name it "correctly" than they were to select foil words that were iconically opposite (e.g., *keex*; $b = -0.40$, $SE = 0.06$, $z = -6.533$, $p < .001$), unrelated (e.g., *cougzer*; $b = -0.28$, $SE = 0.06$, $z = -4.696$, $p < .001$), and even those that were associated with some of the same properties (e.g., *bomburg*; $b = -0.28$, $SE = .06$, $z = -4.635$, $p < .001$; the results are broken down by each individual word in the Appendix). Categorizing words into discrete categories like "similar" and "unrelated" makes for a simpler analysis, but a more powerful way of understanding how iconic relationships influenced people's choices is to predict people's responses from Euclidean distances in the meaning space established by the dimensions for which we had ratings (largeness, roundness, dominance, etc., as described in Section 2.3.1). For example, given a drawing elicited by the word "horgous," we can compute the distance in meaning-space for each word choice ("cougzer," "keex," etc.) and use these (z-scored) distances as a predictor of people's choices in a logistic mixed effects model in place of the categorical *word type* predictor. Higher distances reflect greater *dis*similarity.

As shown in Fig. 3B, distance was a significant predictor of name choice ($b = -0.14$, $SE = 0.02$, $z = -6.50$, $p < .001$), suggesting that while our categorical definition of iconic similarity was a poor fit, iconic similarity nevertheless affected name choices. The greater the distance in meaning-space between the correct word and a foil, the less likely that foil was to be selected as a name for the creature. Overall, the results from the naming task suggest that iconic associations are used to select a name for a never-before-seen creature.

3.2. Rating task

In the naming task, participants tended to select the word that was actually used to elicit the drawing as the best name for the creature depicted in the drawing. This was true even though they had no prior knowledge or experience with the words or the creatures. This result suggests that the words used to elicit the drawings systematically influenced the appearance of the creatures. But in what way? We next examined the relationship between the properties of the words (as rated by participants in the word rating task) and the properties of the drawings.

3.2.1. Properties from the infer a meaning task

In the first set of models, we tested whether drawings elicited by words carrying sound-symbolic properties for roundness, spikiness, largeness, smallness, masculinity, and femininity elicited drawings exhibiting those characteristics. We present the results of each opposing pair, one at a time.

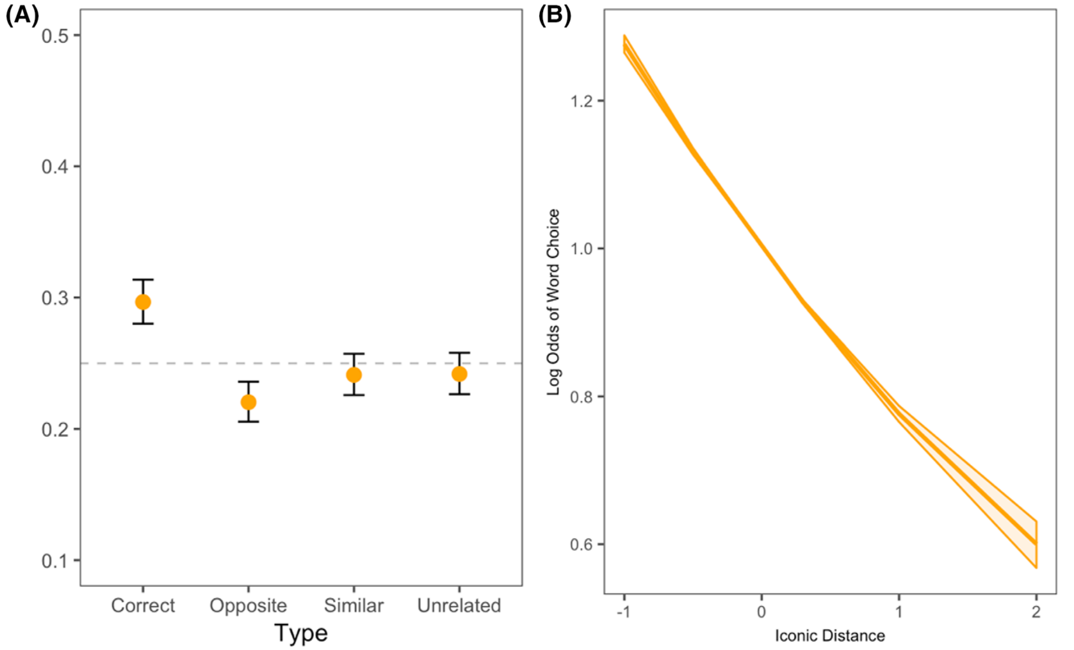

Fig. 3. (A) Results for the recover creature names task showing the proportion of choices by word type aggregated across all of the nonce words. The dotted line shows chance (25%). (B) The predicted probability of a name choice as a function of iconic similarity between a given option word (e.g., cougzer) and the word actually used to elicit the drawing (the correct word, e.g., horgous). Whiskers in A and ribbons in B show 95% confidence intervals (all effects are plotted in R using the packages *effects*; Fox, 2003, and *ggplot2*; Wickham, 2016).

3.2.1.1. Size: The upper row of Fig. 4 shows that words that were rated as sounding *large* (b = 0.25, SE = 0.038, z = 6.68, p < .001) and those rated as sounding *spiky* (b = –0.17, SE = .037, z = –4.61, p < .001) elicited drawings that were rated as larger (i.e., note that we were interested in the perceived size of the drawn creature rather than the size it occupied on the drawing canvas). Words that were rated as sounding *small* (b = –0.23, SE = 0.063, z = –3.71, p < .001) and those rated as sounding *round* (b = 0.13, SE = 0.062, z = 2.15, p = .032) were more likely to elicit drawings that appeared small.

3.2.1.2. Roundness/spikiness: The second row of Fig. 4 shows that words rated as sounding *round* elicited drawings that were judged to be rounder (b = 0.11, SE = .046, z = 2.40, p = .016). Words that rated as sounding *small* (b = –0.13, SE = 0.037, z = –3.36, p < .001) and those rated as sounding *masculine* (b = 0.07, SE = 0.031, z = 2.38, p = .017) elicited drawings that were judged as more spiky. It is perhaps surprising that words rated as sounding *spiky* did not elicit drawings that appeared spiky—however, as shown in Table 1, scores for roundness/spikiness and size were highly correlated, introducing moderate multicollinearity into the model (VIFs = 1.33–2.74). When *size* was removed from the model, spikiness became a significant predictor (p = .01).

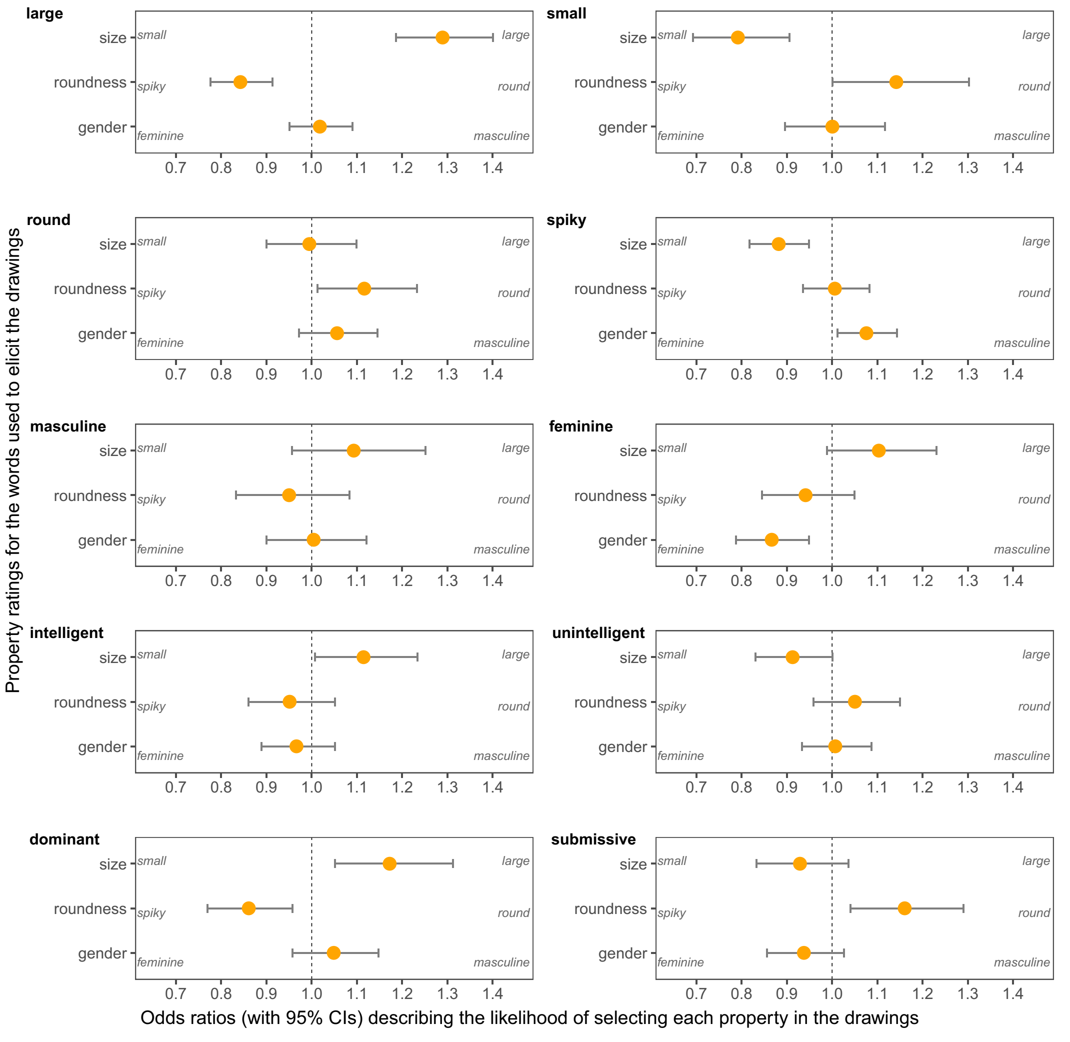

Fig. 4. Results for the infer creature properties task showing odds ratios (ORs) with 95% confidence intervals (CIs) that a given property was selected as a function of the three property rating continua (small–large, spiky–round, feminine–masculine). Each panel summarizes a model (described in Section 2.5). The upper six panels show models predicting properties that were measured in the word ratings. The lower four panels show models predicting more abstract properties.

3.2.1.3. Gender: The middle row of Fig. 4 shows that words rated as more masculine did not elicit drawings that were more masculine. Words rated as sounding more feminine elicited drawings that people rated as feminine ($b = -0.14$, $SE = 0.043$, $z = -3.34$, $p < .001$).

3.2.2. *Generalized properties*

We also hypothesized that sound-symbolic properties of the words would *generalize to* and therefore predict the presence of abstract properties in the drawings. Specifically, we tested whether drawings elicited by words carrying sound-symbolic properties for roundness, spikiness, largeness, and smallness produced drawings carrying abstract properties: Roundness/sharpness might relate to sharpness of intellect, and therefore, intelligence, while the large/small dimension might index physical intimidation (as in Auracher, 2017), and therefore, dominance or submissiveness. These predictions were partially supported.

3.2.2.1. Intelligence: The penultimate row of Fig. 4 shows that words rated as sounding *large* tended to elicit drawings that were rated as appearing intelligent ($b = 0.11$, $SE = 0.047$, $z = 2.31$, $p = .021$), while words rated as sounding *small* tended to elicit drawings that were rated as appearing unintelligent ($b = -.09$, $SE = .043$, $z = -2.12$, $p = .034$). Thus, it seems that *size* associations in the words, as opposed to roundness/spikiness, contributed to ratings of intelligence in the drawings.

3.2.2.2. Dominance: The final row of Fig. 4 shows that words rated as sounding *large* elicited drawings that were rated as appearing dominant ($b = 0.16$, $SE = 0.051$, $z = 3.12$, $p = .002$). This was also true of drawings elicited by words rated as sounding *spiky* ($b = -0.15$, $SE = 0.050$, $z = -2.98$, $p = .003$). Conversely, words rated as sounding *round* elicited drawings that were rated as more submissive ($b = 0.15$, $SE = .051$, $z = 2.95$, $p = .003$).

4. Discussion

The drawings elicited by a given nonce word—a word without a conventional meaning—were similar. Words like “horgous,” rated by other participants as sounding large and round, elicited creatures that appeared larger and rounder. Words like “keex,” rated as small and spiky, elicited creature drawings that were smaller and spikier. Words like “heonia” that were rated as sounding feminine produced drawings of creatures that looked feminine. Asked to draw a creature named a “horgous,” “keex,” and so forth, naïve participants created drawings that could be reliably matched to these labels. Not only were people able to match the drawings back to the words that elicited them, but the pattern of name matches reflected distances in meaning-space between the word that elicited the drawing and the name choices provided. For example, a “horgous” was more likely to be confused with a “bomburg” than with a “keex.”

There was some evidence that word-form iconicity generalized to abstract properties; for example, “larger” words elicited more dominant creature drawings, offering a partial confirmation of our hypothesis that concrete attributes would generalize to abstract properties. Why were *large*-sounding words more likely to produce *dominant*-looking drawings, respectively, while *spiky*-sounding (i.e., sharp) words were no more likely to produce *intelligent*-looking drawings? One possibility is that while size may be a reliable cue to threat, the relation between sharpness and intellect is more symbolic (and perhaps culture-specific). The relationship between iconicity and abstraction is a ripe topic for future work.

Taken together, these results point to the generality with which people draw on form–meaning resemblances even in open-ended situations like our "draw a creature" task. The facility with which people include similar elements in drawings elicited by a given nonce word is difficult to reconcile with views that paint iconicity as a marginal and incidental feature of spoken language (Hockett, 1978; Pinker & Bloom, 1990). However, because the words we used to elicit drawings were selected on the basis of being non-randomly matched to properties, our work leaves open the question of how likely it is that a random nonce word would elicit drawings with similar properties.

In combination with other recent work (e.g., Auracher, 2017; Perlman et al., 2015; Perlman & Lupyan, 2018; Perry et al., 2018; Westbury et al., 2018), we hope that the present investigation helps to move us beyond the simple question of *whether* people make use of non-arbitrary relationships in spoken language (they do!) and toward understanding why natural languages are not even *more* iconic (Lupyan & Winter, 2018; Monaghan et al., 2014) and precisely where the form–meaning associations used by our participants came from.

Author contributions

All authors contributed to the study design. CPD and GL wrote the manuscript. GL collected the data. CPD analyzed the data with supervision from GL. HMM provided critical comments on the manuscript. All authors approve the final version of the manuscript. This material is based upon work supported by the NSF GRFP under Grant No. DGE 1247393 to HMM as well as NSF-PAC 1734260 to GL.

Notes

1. The most cited example is the mil/mal contrast, but Sapir's (1929) report includes several tables of results showing summaries of other contrasts. Unfortunately, this was the age before open data and hence "It would be quite impossible to report on all the details of the experiment in this place" (p. 230).
2. Indeed, the relative shortness of "whale" compared to "micro-organism" is used by Hockett to illustrate the principle of arbitrariness (Hockett, 1978).
3. For further reading on such cross-modal correspondences, we direct the reader to the work of Charles Spence and colleagues (e.g., Deroy, Crisinel, & Spence, 2013; Gallace & Spence, 2006; Parise & Spence, 2012; Spence, 2011).
4. The *Baby Shark* song can be found at https://www.youtube.com/watch?v=XqZsoesa55w (this is not an archival link and may cease working).

5. A useful variant of this design which we did not test would have participants choosing from among all 12 words or include a free response option.

Open Research Badges

This article has been awarded Open Materials and Open Data badges. All materials and data are publicly accessible via the Open Science Framework at https://osf.io/8juyc/. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 298–348.

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*(1), 16–48.

Auracher, J. (2017). Sound iconicity of abstract concepts: Place of articulation is implicitly associated with abstract concepts of size and social dominance. *PLoS ONE*, *12*(11), 1–25. https://doi.org/10.1371/journal.pone.0187196.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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*, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research*, *81*(1), 119–130.

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, *41*(8), 2149–2169.

de Saussure, F. (1959). *Course in general linguistics*. New York: Philosophical Library.

Deroy, O., Crisinel, A. S., & Spence, C. (2013). Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychonomic Bulletin & Review*, *20*, 878–896. https://doi.org/10.3758/s13423-013-0397-0.

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

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech*, *57*(3), 367–393.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of StatisticalSoftware*, *8*(15), 1–27.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203. https://doi.org/10.3758/BF03193720

Hockett, C. F. (1978). In search of Jove's brow. *American Speech*, *53*(4), 243–313. https://doi.org/10.2307/455140

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

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

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*(3), 575–586. https://doi.org/10.1111/j.1551-6709.2010.01169.x

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

Lupyan, G., & Casasanto, D. (2012). The meaning of nonsense words. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *Proceedings of the 9th International Conference on the Evolution of Language* (pp. 490–491). Kyoto, Japan: World Scientific.

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193. https://doi.org/10.1017/langcog.2014.21

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

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322. https://doi.org/10.1111/j.1467-7687.2006.00495.x

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

Newman, S. N. (1931). Further experiments in phonetic symbolism. *American Journal of Psychology*, *45*, 53–75.

Nielsen, A., & 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*(2), 115.

Nielsen, A., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, *67*(2), 153.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325–347). Cambridge, UK: Cambridge University Press.

Parise, C. V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research*, *214*(3), 373–380. https://doi.org/10.1007/s00221-011-2836-3

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3–4), 319–333. https://doi.org/10.1007/s00221-012-3140-6

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*(11), 1419–1421. https://doi.org/10.1177/0956797611421791

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*(8), 150152. https://doi.org/10.1098/rsos.150152

Perlman, M., & Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific Reports*, *8*, 1–14. https://doi.org/10.1038/s41598-018-20961-6

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 1–15. https://doi.org/10.3389/fpsyg.2010.00227

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., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, *21*(3), 1–8. https://doi.org/10.1111/desc.12572

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–727. https://doi.org/10.1017/S0140525X00081061

Plato. (1999). *Cratylus* (C. D. C. Reeve, Trans.). Indianapolis, IN: Hackett.

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

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

Sidhu, D. M., & Pexman, P. M. (2015). What's in a name? Sound symbolism and gender in first names. *PLoS ONE*, *10*(5), 1–22. https://doi.org/10.1371/journal.pone.0126809

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

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, *112*(19), 5968–5973. https://doi.org/10.1073/pnas.1423080112

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology*, *64*(12), 2392–2404.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language, Vol. 2: Phonology* (pp. 527–568). Stanford, CA: Stanford University Press.

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

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Appendix

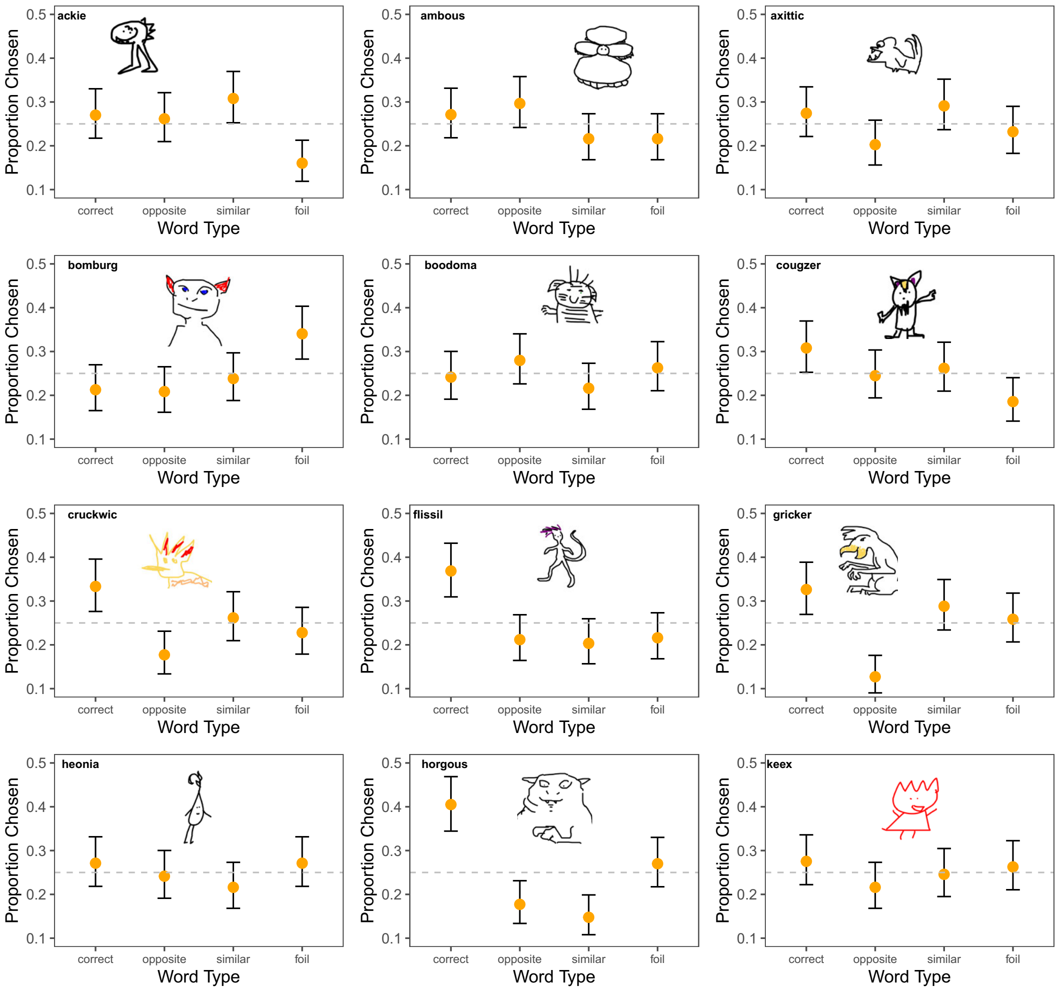

Fig. A1. The proportion of choices by word type for each word in the "recover creature name" task (analogous to Fig. 3). A representative drawing of each word is placed in the plotting area. In each panel, the dotted line shows chance (25%).