Distribution Agreement

Signature:

______________________________ ___________________

Leonardo Michelini Date

Is sound symbolism an affordance?

By

Leonardo Michelini

Master of Arts

Psychology

Lynne C. Nygaard, Ph.D.
Advisor

Stella Lourenco, Ph.D.
Committee Member

Phillip Wolff, Ph.D.
Committee Member

Accepted:

Kimberly Jacob Arriola
Dean of the James T. Laney School of Graduate Studies

Date

Is Sound Symbolism an Affordance?

By

Leonardo Michelini

B.S., Georgia State University, 2019

Advisor: Lynne C. Nygaard, Ph.D.

An abstract of a thesis submitted to the Faculty of the James T. Laney School of

Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology

2022

Abstract

Is Sound Symbolism an Affordance?

By Leonardo Michelini

Sound symbolism is a non-arbitrary link between word and meaning. Over the last decades, studies have emphasized the ubiquity of this phenomenon in language and argued that it is an important factor in understanding language evolution and development. However, precisely characterizing the role that sound symbolism plays and how it is situated within the broader communicative system has been a difficult task. Researchers have tentatively proposed that sound symbolism can be understood as an affordance, but this notion remains underexplored. Thus, this investigation is chiefly concerned with fleshing out this hypothesis and testing it empirically. If sound symbolism is understood as a communicative tool that facilitates linguistic and perceptual processing, this functional purpose could be interpreted as an affordance that constrains the range of sound-to-meaning mappings. Alternatively, sound symbolism itself could be considered an affordance for iconic representation in prosody (e.g., depicting the size of the referent using tone of voice). A preliminary experiment was conducted to determine whether our materials, animal images and pseudowords, evoked the predicted size magnitude. A second experiment assessed whether participants would assign labels to complex objects based on shared size similarities between the two. Finally, the third experiment examined whether sound-symbolic word-referent pairs would elicit stronger prosodic modulation compared to pairs that were not. We found evidence that sound-to-meaning mappings were driven by affordances in experiment 2. However, the results from experiment 3 were largely inconclusive, and thus the status of the affordance hypothesis in the realm of prosody is still uncertain.

Is Sound Symbolism an Affordance?

By

Leonardo Michelini

B.S., Georgia State University, 2019

Advisor: Lynne C. Nygaard, Ph.D.

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of

Emory University in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology

2022

Acknowledgements

First and foremost, I am grateful for my advisor, Lynne Nygaard, who has guided me from the very inception of this research project through her wisdom, knowledge, and kindness. She has, among many other things, helped think about sound symbolism and language, design studies, analyze data, and also provided her time and labor to revise this thesis. I am also indebted to my committee members, Phillip Wolff and Stella Lourenco, for their invaluable feedback.

I would also like to thank the members of the Speech and Language Perception Lab. I thank Ana Maria Hoffman, for our riveting theoretical discussions, and reassurance when I felt completing this thesis was an insurmountable task. I thank my undergraduate collaborators, Abigail Stamper and Hyemin Yoon, for helping implement the experiments and aiding in the gargantuan task of preprocessing the auditory data. I also thank former post-doctorate researcher Christina Tzeng for her assistance with auditory analysis.

Finally, this research project was only made possible with the love and support of my family; especially my parents, Ana and Marcelo, and my sister, Luiza, who have always encouraged me in my intellectual pursuits, and my wife Viviane, who has dutifully proofread my writings and listened to me ramble on about boubas and kikis.

Table of Contents

Is Sound Symbolism an Affordance?

Introduction

Word-referent relationships in language have traditionally been understood as arbitrary, that is, mediated entirely by social convention (Hockett, 1960; Saussure, 1983). According to this view, the only reason that the symbol “tree” refers to the real-world object of a tree in English is that a speech community agreed to call it that. However, over the last decades researchers have displayed renewed interest in the phenomenon of sound symbolism, which undermines the theoretical postulate of the “arbitrariness of the sign” (Saussure, 1983). Sound symbolism is a broad umbrella term defined as any link between word and meaning that is motivated instead of arbitrary (Dingemanse et al., 2016). An oft studied case of this is iconicity. When a sign bears resemblance to what it represents, this is a relationship of iconicity. A map represents the topography of the geographical space that it charts, and a smile emoji, in a stylized fashion, looks like a person smiling.

The emblematic example of iconicity in the context of language is onomatopoeia. These are performative, phonologically unusual words that imitate the sounds they describe. Animal noises like a rooster’s crow are frequently labelled with onomatopoeic words in various languages: cock-a-doodle-doo, cocoricó, ku-kudu-koo, ü-ürü-üüü (English, Portuguese, Hindi, and Turkish, respectively). The specific phonology of these words is superficially distinct because they are adapted to the rules and conventions of each language, but their form is constrained by the properties of the sounds they resemble. This results in cross-linguistic similarity; especially when comparing onomatopoeia to other, less iconic words. In this example, all the different words that emulate a rooster’s crow encode a specific stress pattern (roughly, strong-pause-weak-weak-strong).

Historically, language researchers have acknowledged these counterexamples to the arbitrariness principle like onomatopoeia, but dismissed them as marginal cases that do not represent the general rules of language (Hocket, 1960; Saussure, 1983). However, recent studies have shown that iconicity is shot through the lexicon (Dingemanse, 2018). Certain phonemes are more likely to refer to specific meanings across languages (e.g., nasals like /n/ are used in words for "nose"), and within languages perceptual properties can predict the sound structure of words (Wichmann et al., 2010; Blasi et al., 2016; Sidhu et al., 2021; Winter & Perlman, 2021). These findings cannot be accounted for if the sounds that constitute words are exclusively selected by social convention. Therefore, the very notion of what language is has to be revised to accommodate the pervasiveness of iconicity (Perniss et al., 2010).

In addition to what it implies for the nature of language, sound symbolism might also be an important factor for understanding language learning and evolution. One of the advantages of iconicity for communication is that it makes messages easier to understand in the absence of a shared conventionalized system. A drawing of a stick figure can be interpreted by members of diverse populations, even if they do not share the same language. Accordingly, studies have demonstrated that participants from different countries reliably choose the same referents for non-lexical vocalizations (Perlman et al., 2022; Silva et al., 2020), and quite a few experiments show that, using cues from phonetic properties, individuals can guess the meaning of contrastive terms like hot and cold in languages they do not speak (Tsuru & Fries, 1933; Brown et al., 1955; Klank et al., 1971; Kunihira, 1971; Lockwood et al., 2016; Tzeng et al., 2017).

While this is not direct evidence of what a hominid proto-language may have sounded like, it suggests that iconic signs may have served as a foundation upon which a more complex system of communication was built. For example, one of the key features of words is that they are able to call upon displaced objects; they can refer to things that are not present in the immediate environment. How would this ability emerge in a proto-language based on iconicity? A speaker could simply imitate features of an object, and the listener, in turn, would be able to pick out the displaced referent because the shared properties between the vocalization and the object would evoke perceptual representations in the mind of the listener (Imai & Kita, 2014; Perniss & Vigliocco, 2014). As we have seen, iconic depiction makes this possible even without a shared agreement regarding what each vocalization means. Through a process of ritualization (i.e., use outside the original context), vocalizations that were once used when trying to pragmatically communicate an idea on one occasion could find use in more and more situations, and gradually become conventionalized.

Researchers have also proposed that iconicity serves a similar role in scaffolding language learning in ontogeny (Imai & Kita, 2014). Sound-symbolic words are used more often by infants and by their caregivers when speaking to them (Perry et al., 2018), and this has beneficial effects on word learning (Imai et al., 2015; see Tzeng et al., 2017 for evidence suggesting this might only happen later in development than in infancy). Because iconic depiction is intuitively meaningful, it can bridge the gap between imitation and more complex forms of representation. In this way, children might learn the crucial "referential insight": understanding of the idea that symbols can stand for objects, places, people, etc. (Imai & Kita, 2014; Perniss & Vigliocco, 2014; Nielsen & Dingemanse, 2020).

The protean nature of iconicity

A central difficulty in conceptualizing sound symbolism is that it seems to operate in parallel with the abstract symbol representation described by the traditional view. The sounds in “balloon” express sensory information: liquids like /l/ are associated with roundedness and are therefore more likely to be present in words related to this feature (Sidhu et al., 2021). Yet, the label “balloon” also strictly denotes a precise class of objects, and not just a vague and porous category of “things that are round.”

Further, speech sounds must be multiuse tools. Languages have a limited phonemic inventory that could not possibly express as many meanings as iconicity can represent if sounds had singularly defined mappings. For example, voiced consonants such as /b/ are associated both with large size and round shape (Westbury et al., 2018; Sidhu et al., 2021). Consequently, sound-to-meaning mappings must be pluripotential: speech sounds relate iconically to any number of objects depending on the situation (Winter et al., 2021). This does not mean that iconic signs are subject to duality of patterning (Hockett, 1960). Properly arbitrary symbols can, in theory, be recombined to generate an infinite number of meanings that are discriminable. The difference between “cat” and “mat” is only one phoneme, but these are completely distinct objects. Iconic signs, by contrast, need to resemble their target meaning; onomatopoeias for different bird songs are likely to be similar to each other. But the same sounds that describe animal noises must also be capable of iconically representing different perceptual features, motion, psychological states etc. Thus, sound symbolism must be a malleable and context sensitive semiotic device. In fact, this might explain why many findings in this literature fail to replicate and why sometimes effects reverse direction when the experimental designs suffers minor alterations (Winter et al., 2021). The challenge, then, is determining what constrains the range of possibilities for sound-symbolic expression.

Affordances

In response to these issues, some researchers have suggested that we conceptualize sound symbolism as an affordance that linguistic symbols offer to create meaning (Dingemanse et al., 2016; Occhino et al., 2017). Affordances are relational properties that naturally arise from basic characteristics of objects in the environment and basic characteristics of organisms interacting with them (Gibson, 1979). Essentially, affordances are what an animal can *do* with a thing. A hard, smooth, horizontal surface connected to a vertical surface—a chair—affords sitting for us humans because of how our skeletons are built and where our joints fold. For a deer, a chair does not invite sitting. While linking sound symbolism to affordance seems like a promising proposal, it is unclear precisely how these ideas fit together. In the following sections, I will present three mutually compatible formulations of the hypothesis that sound symbolism is an affordance.

Affordances of the articulators

Language researchers until the mid-20th century considered vocal transmission to be a constitutive feature of language (Hockett, 1960; Saussure, 1983). Nowadays, scientists recognize that, while sound is the most common conduit for language, it can be instantiated in a variety of modalities like hand signs or touch (Goldin-Meadow & Brentari, 2017). The physical makeup of symbols has critical implications for the expression of iconicity. Comparing spoken and signed languages reveals that some semantic domains are rated as more iconic in the former, and some are rated as more iconic in the latter (Perry et al., 2015). Words related to body parts, artifacts, and numbers highly iconic in sign languages, but not in spoken languages (Thompson et al., 2020). Conversely, words that refer to color and sounds in spoken languages are significantly more iconic than in sign languages (Perlman et al., 2018). Furthermore, beyond iconicity, signers can also refer to objects via spatial deixis; pointing to objects conveys the idea that you are talking about this or that thing. This opens the door for a vast range of non-arbitrary links to meaning; namely, indexical signs.

These findings all conform to the view that the articulators of language—vocal tract, hands—afford fundamentally different forms of non-arbitrary symbolic expression (Perlman & Cain, 2014; Dingemanse et al., 2015). The possibility of pointing and simulating motor programs that hand signs afford appears to make iconic depiction more likely in some semantic domains and unlikely in others. Likewise, the affordances that sound offers, like representing intensity using pitch, make iconicity more prevalent in some areas and less so in others. This version of the affordance hypothesis is well defined conceptually and empirically supported. Thus, I will focus on less clear-cut versions of this basic idea.

Affordances of word-referent pairs

Researchers have suggested that iconic sound symbolism is based on general cognitive mechanisms such as cross-modal correspondence, which lends them, to some extent, the ability to transcend language barriers (Sidhu & Pexman, 2017). However, as discussed earlier, there are infinitely more meanings that we can express iconically compared to the number of distinct speech sounds in any given language. Therefore, while cross-modal correspondences might provide us with a range of possible basic perceptual mappings, as a phenomenon of language, sound-symbolic representation itself must be the result on an active link forged by an interpreting organism that recognizes shared characteristics between speech sounds and certain objects and features (Occhino et al., 2017). The question, then, becomes when and why are these connections established. This is where the idea of affordances becomes useful because they imply agency and interpretation on the part of an organism, but the specific actions afforded by objects are nonetheless predictable based on characteristics of the two parties. Are there characteristics of speech sounds and referents that, when put together, create affordances for iconic representation? Evidently, they need to share features, but what exactly this entails is complicated. If a word that describes a large round object has a /b/ sound in it, is this because it depicts its big size or its curvilinear shape? The answer might be related to the function that sound symbolism fulfills.

Language comprehension occurs in a remarkably short time span. Within this small window, we must process vast amounts of information, 10-15 phonemes a second, because after about 100 milliseconds the memory trace for the speech stream has already deteriorated. We must not only perceive the speech sounds correctly but interpret their semantics and embed them within a sentence structure. These challenges in language processing have been called the "Now-or-Never Bottleneck" (Christiansen & Chater, 2015). Researchers have hypothesized that we use many strategies to solve this problem, and they might be useful in explaining when we use sound symbolism. Linguistic information is simultaneously encoded at different levels of hierarchically structured representations; for example, at once as sounds and as words (Christiansen & Chater, 2015). This occurs through a process of chunk-and-pass processing. That is, information is rapidly condensed into a meaningful unit and passed along to systems that deal with more abstract representations.

This model explains the dual role that sound-symbolic words play; at the higher levels of the hierarchy, it is a proper arbitrary symbol, but at the lower level its sounds are cross-modally matched to their referent. Congruency between sound and meaning early in the chain could facilitate later process by serving as a cue for predictive processing later down the line (Clark, 2013). For example, vocalizations that evoke disgust (e.g., icky) could prime the semantic context of the word and make the interpretation of its semantic meaning easier. In fact, studies using speeded classification paradigms and implicit association tasks showed that congruent mappings between sign and meaning were processed faster than incongruent ones (Westbury, 2005; Kovic et al., 2010; Parise & Spence, 2012; Ohtake & Haryu, 2013). Furthermore, the effect runs in the other direction as well: presenting individuals with pseudowords allowed them to perceive masked visual information faster when the sound-to-meaning mapping was congruent (Hung, Styles, & Hsieh, 2017; Heyman et al., 2019).

Returning to the idea of affordances, the structure of words may invite particular iconic mappings when there is an inherent similarity between sound and referent, but also the use of iconicity facilitates prediction in perceptual or linguistic processing. For that to be the case, iconicity would have to be associated with particularly salient features of the referent that help distinguish it from other objects or other classes of objects. As an illustrative example, in the Jívaro language spoken by indigenous Amazonians, the sounds in animal names are correlated with size magnitude (Berlin, 2006). The words for smaller species of fish and birds tended to have high-front vowels such as /i/ in them, whereas then words for bigger animals tended to have low-back vowels such as /u/. These are examples where sound symbolism is facilitating within-class distinction; it is indexing different types of birds, which are members of a shared group. Presumably, the context of “bird” affords different iconic mappings than a context like “all vertebrates” would, because size would not be particularly diagnostic of a species of bird when comparing it to apes, snakes, sharks, etc. When the frame is specific enough like with “bird”, referents can be reliably picked out by marking them according to their size differences, whereas if one was comparing a bird to every other vertebrate, another mapping that was maximally discriminative, that helped tell this particular species of bird apart from all the other vertebrates, would have to be selected instead.

Affordances of the multimodal communicative system

Phonemes are defined by spectral properties. For example, one difference between an /i/ and an /e/ is that the latter has a higher first formant (F1) frequency, which is inversely correlated to vowel height (Abercrombie, 1982). To produce /e/, one must open their vocal tract, and thus lower the tongue further than when they are produce /i/. As such, /e/ is a lower vowel than /i/. We can also alter the fundamental frequency (F0) of our productions, which is the lowest frequency at which our vocal cords vibrate at a given time and is perceived as pitch. F0 does not determine phoneme quality; higher or lower pitch does not dictate whether a sound is perceived as, say, a low or high vowel. Nonetheless, though we can modulate F0 independently of which phoneme is being produced, we tend to pronounce vowels at different pitches, and consonants alter the pitch of subsequent vowels (House & Fairbanks, 1953; Hillenbrand et al., 1995).

Recently, researchers have been calling attention to the multimodal nature of human communication (Levinson & Holler, 2014; Vigliocco et al., 2014). Although language can be conceptualized as a system of rules that govern symbol combination, in everyday life it is overwhelmingly used as a tool to communicate interpersonally, and is highly integrated with other modalities like prosody, gesture, and bodily expression. And it is not only phonology that can be used for iconic depiction; prosody is frequently used in this way too (Nygaard et al., 2009; Herold et al., 2011; Silva et al., 2020). For example, we can use pitch in speech to express the brightness level of colors (Tzeng et al., 2017; Tzeng et al., 2019). But do we employ these different channels of iconic representation in a reinforcing or complementary fashion?

Cwiek and colleagues (2021) have suggested that the expressive affordances of phonemes might lead speakers to pronounce words a certain way consistently to maximize their perceptual salience. Imagine the following scenario: a word that refers to a large object contains a mid-vowel. Speakers who want to accentuate the size of the referent can pronounce the vowel with lower spectral properties than the standard mid-vowel to exacerbate its sound-to-meaning link. Over time and repeated use, the vowel quality becomes lower, and this becomes the standard form. If we take this macroscale language evolution model and apply it to a smaller timeframe, it implies that individuals are more likely to use prosody as an expressive tool when the phonology of a word already contains latent iconic mappings. That is, sound-symbolic word form affords reinforcing iconic representation in prosody. For example, this hypothesis would predict that someone trying to emphasize the small size of an object might use terms that are already sound-symbolic such as "teeny tiny" or "itty bitty", but then further reinforce this perceptual feature by speaking with a high-pitched voice. This could facilitate language processing, as described above, by redundantly hinting at a semantic space at two different levels of representation: phonology and prosody (Christiansen & Chater, 2015).

However, possibly in opposition to this hypothesis, Tzeng and colleagues (2019) proposed that individuals use prosody to communicate referential information when words fail to distinguish between objects. When participants referred to colors using distinct words, like "blue" and "purple", they did not modulate their tone of voice, rate of speech, or intensity. But when they talked about objects that could be described with the same label, like different kinds of red both referred to simply as "red", they used these prosodic features to communicate different brightness levels. This suggests that individuals use prosody as a disambiguating "vocal gesture" in case other channels of communication underspecify reference (Tzeng et al., 2017, 2019). Though this study did not directly investigate how prosody interacts with sound symbolism, it could imply that iconic depiction is employed in prosody precisely when word form does not.

The present study

Size-sound symbolism is one of the best understood instances of sound-to-meaning mapping. The first investigation on this topic dates back all the way to Sapir (1929). Since then, a litany of studies has established that speech sounds are reliably matched to size magnitude (Ohala, 1994; Berlin, 1994; Diffloth, 1994; Klink, 2000; Berlin, 2006; Tsur, 2006; Shinohara & Kawahara, 2010; Blasi et al., 2016; Westbury et al., 2018; Godoy et al., 2019). High-front vowels and stop consonants are associated with smallness. Low-back vowels and voiced consonants are associated with largeness. The frequency code hypothesis is an influential model that tries to tie these findings to a broader evolutionary context (Ohala, 1994). These judgments about size are explained by differences in frequency. Specifically, low energy in fundamental and formant frequencies is associated with large size, and high energy in fundamental and formant frequencies is associated with small size. Ohala (1994) further claimed that these mappings are phylogenetically ingrained and that they are observed in animal communication, wherein low frequency indicates dominance and aggressiveness, and high frequency signifies submissiveness.

Laboratory experiments investigating size-sound symbolism were often based on a pseudoword-referent matching paradigm. Participants are shown an image and asked to assign a pseudoword to it. However, many studies included a small number of pseudowords and presented participants with a binary choice (e.g., “is this small table a ‘mil’ or a ‘mal?’”). Under these conditions, the research question could become transparent and introduce demand characteristics. Participants might pick up on what the researchers are assessing and provide responses that conform to their expectations. To address some of these issues, Thompson and Estes (2011) designed an experiment that had objects (greebles) of multiple sizes and pseudowords with varying increments of big-sounding phonemes. They demonstrated that sound-symbolic associations could be made in a graded function: the larger the object, the more big-sounding phonemes the pseudoword contained.

Nonetheless, Thompson and Estes (2011) used simplistic visual stimuli that only differed in size. But objects of language are frequently complex and multidimensional, and different in many aspects besides size, which calls into question the applicability of these findings to natural language. Additionally, to understand what factors afford iconic representation, this type of design is inadequate because there is only one possible sound-to-meaning mapping available. Thus, in all our tasks, we included a large inventory of pseudowords with varying amounts of small or big-sounding phonemes, and images that differed in respect to multiple features. If sound-symbolic mappings occurred in this case, it would not be because they were the only options possible given the stimuli, but because they were afforded by the communicative advantage they provided. Namely, that they would be informative in distinguishing between the different referents.

Experiment 1a examined whether our materials were perceived as expected; that our small, medium, and big pseudowords were judged as such. Experiment 1b determined whether size was the most salient characteristic of our images, considering that they were multidimensional. Following up on this, experiment 2 assessed which labels participants picked for these complex referents.

To investigate how iconicity in word form and prosody would interact, in experiment 3 we presented participants with materials from experiments 1a and 1b, but preassigned them to specific referents. Participants were instructed to produce a carrier sentence followed by a target pseudoword. Our basic assumption was that if sound symbolism in the word form affords further iconic depiction in prosody, when participants pronounced names that were congruent with the image (e.g., big-sounding pseudoword with a big object), they would be more prone to using prosody to represent the size of the object.

Every study described below was created and hosted in Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2018). Participants were recruited using the crowdworking platform Prolific (https://www.prolific.co). All statistical analyses were performed with R Studio version 4.0.3 (R Core Team, 2020).

Experiments 1a and 1b

These experiments were performed to establish that our materials appropriately represented size magnitude. Participants were asked for their explicit judgments regarding the stimuli. Based on prior findings, we generated an inventory of pseudowords that were hypothesized to vary according to their perceived size (Sapir, 1929, Ohala, 1994; Berlin, 1994; Diffloth, 1994; Klink, 2000; Berlin, 2006; Tsur, 2006; Shinohara & Kawahara, 2010; Blasi et al., 2016; Westbury et al., 2018; Godoy et al., 2019). We expected that high-front vowels and voiceless stops would be interpreted as small and back vowels and voiced stops as big. Pseudowords were exclusively composed from small-sounding phonemes, exclusively composed from big-sounding phonemes, or they mixed-in both. Thus, we predicted that the pseudowords that only incorporated “small” sounds would be rated as the smallest, the pseudowords that only had “big” sounds would be rated as the biggest, and the pseudowords that blended “small” and “big” sounds would be rated in-between the other two types of pseudoword.

For our visual stimuli, we used animal silhouettes. Three distinct creatures belonging to a shared category (e.g., dogs) were always presented together, and each had a different size. Only one exemplar was the target stimulus in any given trial. Because our images were multidimensional, we asked participants to rate them according to a variety of features instead of just size. The objective was determining whether size would be the most diagnostic characteristic. This was operationalized as difference across individuals within each class of animal; if size helped tell them apart, the contrast in size ratings between them would be bigger than in any other dimension.

Participants

Participants were all naive monolingual English speakers with no history of speech or hearing disorders. Thirty participants (mean age = 22.95, SD = 3.53, 2 men, 1 non-binary) rated the pseudowords and were awarded US$ 2.65 (average of US$ 7.95/hour) for participation[1]. Forty participants (mean age = 26, SD = 5.40, 19 men, 1 non-binary) rated the animal images and were compensated US$ 2.75 (average of US$ 7.28/hour).

Stimuli

Pseudowords. Front-high vowels /i, I, e/ and stop consonants /t, k, p/ were selected as the “small” elements, and the back vowels /a, o, u/ and voiced consonants /d, g, b, w/ were selected

[1] On July 24th, 2021, a social media influencer posted a video to Tiktok mentioning *Prolific*. Her followers, mostly young women, joined the platform en masse, which caused a significant gender imbalance and age skew in many studies over the following month. This was one such study. See Charalambides, 2021.

as the “big” elements. Pseudowords all had a CVCV structure, with each syllable being entirely small (e.g., /ti/) or entirely big (e.g., /ba/). As such, small words had two small syllables, and thus only small sounds. Big words had two big syllables, and thus only big sounds. “Medium” pseudowords were created by mixing small and big sounds. There were 2 types of medium words, which either began with a small syllable and ended with a big syllable (henceforth referred to as cvCV medium words) or began with a big syllable and ended with a small syllable (henceforth referred to as CVcv medium words). Given these parameters, we generated a list of pseudowords that spanned all possible permutations. Words ending in /I/ were eliminated to avoid ungrammatical phonology in English. 28 big words were not included due to human error. In total, stimuli included 54 small words, 180 medium words (108 cvCV, 72 CVcv), and 116 big words (see Appendix 1 for a complete list). To ensure that participants would perceive the phonology of the words as intended, they were spelled so there would be little ambiguity regarding their pronunciation (e.g., “peekee” instead of “piki”).

Animal Images. We created 6 categories of animals: rodents, bugs, spiders, birds, dogs, and butterflies. Each category had 3 members, for a total of 18 exemplars. Creatures were represented with PNG images of their silhouettes. All images were downloaded from http://clipart-library.com. Silhouettes were used, rather than drawings or photographs, because they could be transformed with minimal distortion while still retaining identification of different members of a category. Images were imported to Photoshop, where the digital canvas had 4 equidistant, parallel horizontal lines. One member in each category was resized to fit between the first and second lines, one fit the between the first and third lines, and one fit between the first and fourth lines. As such, the smallest animal was consistently one third the height of the biggest, and the middle-sized animal was two thirds the height of the big one (see Picture 1 for an example, or Appendix 2 for the complete set). The small animal was always to the left, the medium in the middle, and the big to the right.

Picture 1. One of the animal categories, butterflies. Every category had 3 exemplars, small, medium, and big, arranged from left to right.

Procedure

Pseudoword ratings. The instructions informed participants that some words in English can evoke sensations or suggest physical qualities. They were asked to read the target pseudoword and rate how big or small it sounded. In every trial, participants saw a single pseudoword at the center of the screen, and at the bottom a Likert scale that ranged from 1 (very small) to 7 (very big). They completed 3 practice trials that were not included in analyses, saw the instructions once again, and then moved on to the experimental trials. Each participant went through a total of 350 trials. The order of presentation of pseudowords was randomized.

Animal image ratings. On each trial, participants saw an image depicting all members of a category of animal. A red arrow pointed to one of the three creatures. Participants were instructed to rate that specific member on a 7-point Likert scale. They rated the animals on 7 different dimensions: size, shape (pointy or rounded), approachability, valence (good or bad), speed, threat, and arousal (adapted from Tzeng et al., 2017). After reading the instructions, participants completed 4 practice trials and saw the instructions again before beginning the experimental trials. Because some participants did not use the full range of the scale in experiment 1a, we added attention trials to ensure data quality. Sometimes there would be an image with 3 animals but no arrow above any of them. In that case, participants had to pick the number 4 on the scale. Trials were divided into 7 blocks corresponding to the dimensions rated. Each block consisted of one trial per each exemplar creature, plus one where there would be no red arrow pointing to the animals, adding to 19 trials per block. Across all the blocks there were 133 trials. The order of the blocks was counterbalanced across participants via Latin Square method, and the order of the trials within each block was fully randomized.

Results and Discussion

Pseudoword ratings. We divided pseudowords into 4 different categories for analyses: small, medium (cvCV), medium (CVcv), and big. To assess the perceived size of each type of pseudoword, we constructed a linear mixed-effects model using the lmer() function from the lme4 package in R (Bates et al., 2015). Likert-scale ratings were considered the predicted variable and the type of pseudoword the fixed effect. The random effects structure consisted of random intercepts for the individual pseudowords, and random intercepts and slopes for participants. 2 participants were excluded from analyses because they used fewer than 5 items on the 7-point Likert scale. A likelihood-ratio test, performed with the mixed() function from the R package afex (Singmann et al., 2021), established that a model including the type of pseudoword as a fixed effect was a significantly better fit for the data than one without it, $\chi^2(3) = 31.71$, $p < .001$.

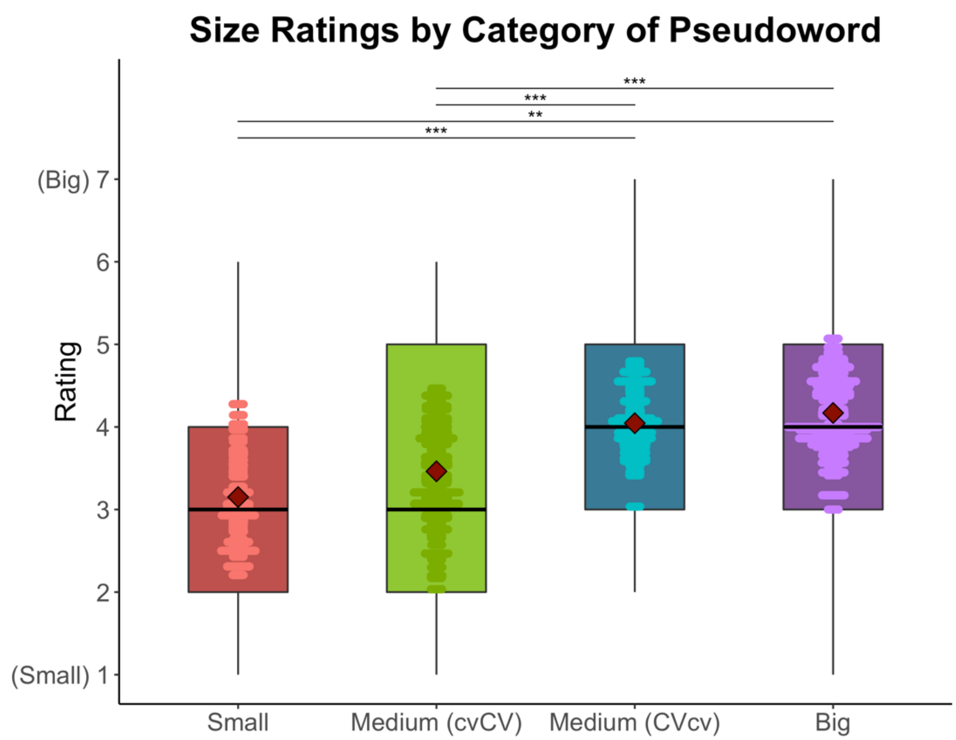


Figure 1. Boxplots compare the distribution of ratings between types of pseudoword. The horizontal lines at the center are the medians, the red diamonds are the means, the colored portions the 0.25 and 0.75 quantiles, and the tails the 0.025 and 0.975 quantiles.

We performed a post-hoc Tukey HSD test correcting for multiple comparisons to acquire the mean rating for each type of pseudoword with the lsmeans() function from the lsmeans R package (Lenth, 2016). The phonemes we hypothesized to be small and big did influence participants' estimations of size. Small words were rated the smallest (M = 3.14 CI[2.80, 3.48]) the big words were rated the biggest (M = 4.11, CI[3.78, 4.45]), and ratings of size for the cvCV and CVcv medium words were intermediate (M = 3.44 CI[3.21, 3.66], M = 4.00, CI[3.77, 4.24]). The order of syllables within medium words also made a significant difference in ratings of perceived size. The CVcv medium words were rated as 0.57 bigger than the cvCV medium words ($p < 0.001$). In fact, it is unclear whether the cvCV and small words or CVcv words and big words were truly distinct given the insignificant p-values in these contrasts ($b = 0.30$, $p = 0.13$: $b = 0.11$, $p = 0.87$). Nonetheless, in line with our predictions, small words were rated as smaller than CVcv and big words (b = 0.87, p < 0.001, b = 0.98, p = 0.002), and cvCV words were rated as smaller than big words (b = 0.68, p < 0.001).

The current methods do not clarify why the two types of medium words differ, or why they were perceived as being so similar to the small and big words depending on nothing more than the order of their syllables. This could be due to a serial position effect, where the first few sounds in a word play an outsized role in how their sound-symbolic content is interpreted. Another possibility is that this was a result of how these pseudowords might be pronounced. They were all disyllabic and had CVCV form. English speakers are wont to stress the first syllable in this case. Thus, this apparent serial position effect might indicate that accent patterns modulate sound-symbolic associations.

One criticism that might be levied against our general interpretation of these results is that, because we presented participants with words on a screen instead of sounds, visual features of the letters themselves may be promoting the iconic mappings. However, previous studies have shown that participants trying to guess the meaning of foreign-language words have similar accuracy rates whether they see graphemes or listen to speech (Tsuru & Fries, 1933; Brown et al., 1955; Klank et al., 1971; Kunihira, 1971; Lockwood et al., 2016; Tzeng et al., 2017). This suggests that people access auditory-based phonological representation when they read and draw sound-symbolic connections from the phonemes, even when presented with printed words.

Further, in a separate task, we asked participants to match a subset of pseudowords to images of differently sized animals, in a similar procedure to the one described below in experiment 2. Participants then completed 4 different assessments of visual and auditory imagery; that is, how vivid their internal representations of sights and sounds were. Using the lm() function from base R (R Core Team, 2020), we regressed scores on the sensory imagery inventories on the rate of "correct" responses in the sound-symbolic matching task (correct responses were coded as a congruent matching such as a "big" pseudoword associated with a big animal). Animal figures were directly shown to the participants, and thus did not need to picture them at all. Thus, visual imagery scores were not related to performance in the sound-symbolic matching task ($b = -0.17$, CI [-0.45, 0.11], $t(26) = -1.23$, $p = .229$). However, the general ability to engage in auditory imagery was, in fact, associated with "correct" mappings between pseudowords and animals ($b = 0.37$, CI [0.03, 0.71], $t(26) = 2.22$, $p = .035$). Because pseudowords were presented in written form, the finding suggests that participants who were better able to imagine what they sounded like were also more prone to making size-sound correspondences. Nonetheless, it should be noted that due to a relatively small number of participants ($n = 29$) and high variability across them, confidence intervals for these regressions are fairly large. Therefore, results from this study should be interpreted with caution. However, given all the evidence presented, it is unlikely that graphemes affected magnitude judgments of pseudowords in experiment 1. But it is possible that presenting written words instead of speech reduces power in this kind of study, as the ability to imagine how pseudowords sound can influence whether sound-symbolic mappings are made or not.

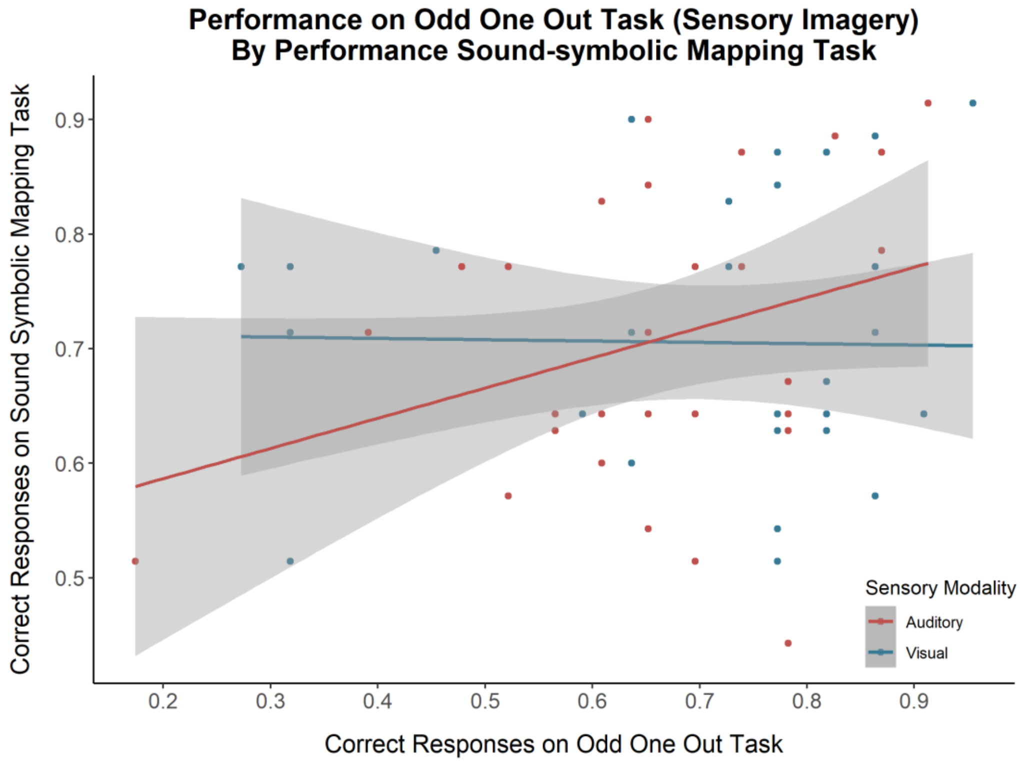
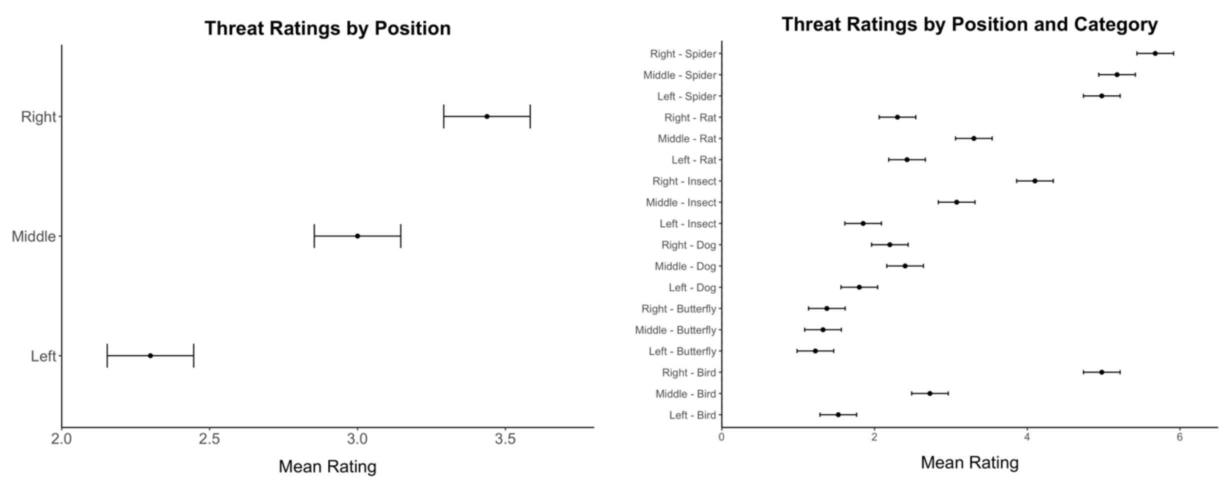


Figure 2. X and y axes index the rate of correct responses (correct responses divided by total trials). Points represent the performance of individual participants.

Animal image ratings. No participant failed the attention trials, and thus all were included in analyses. To determine whether size was the most salient perceptual feature distinguishing the members of each category, we created 7 linear mixed-effects models with the lmer() function from the lme4 R package (Bates et al., 2015). Each model included ratings for one dimension as a predictor. The fixed effects were the category of animal, the position of individual exemplars (coded here as left, middle, and right), and the interaction between these factors. Models with random slopes resulted in singular fit; thus, only random intercepts were computed for participants (see Brown, 2021). If a given dimension indexed differences between the individual members in each category, then position or the interaction between position and category of animal would improve model fit relative to a base model containing only the random effects structure. Likewise, if dimension ratings varied across animal category (e.g., butterfly, insect), then category or the interaction between position and category of animal would be expected to improve model fit.

Comparing models for best fit with log-likelihood tests with the mixed() function from the afex package (Singmann et al., 2021) revealed that the type of animal influenced ratings for all dimensions. For example, there was a significant effect of category on valence ratings ($\chi^2(5) = 527.22$, $p < .001$), which reflects, among other things, that for example participants reported liking dogs and really disliking spiders ($b = 4.18$, $p < .001$). A model including position explained significantly more of the variance than a model without it for all dimensions (all p's > .05) except speed ($p = .051$) and valence ($p = .094$). In all cases, there was a significant interaction between position and category, which indicates that relationships between exemplars varied depending on the creature. For example, though leftmost animals were less threatening and rightmost animals were more threatening than the middle ones, this pattern only held for some of the categories (Figures 3a and 3b).

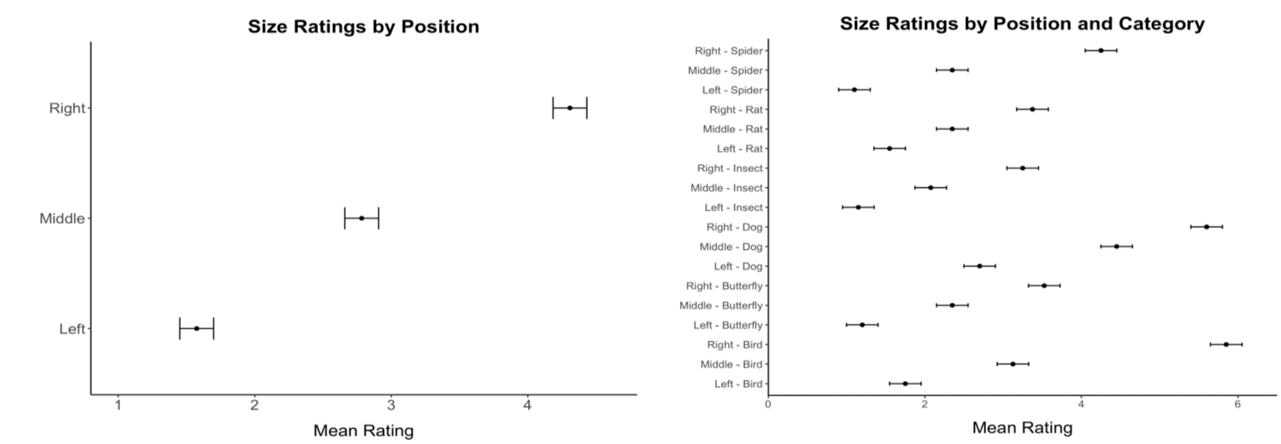


Figures 3a and 3b. a) There is an overall difference in means between threat levels of different exemplars of animal within each class; the leftmost (and thus smaller) creature tends to be less threatening and the rightmost (and thus bigger) creature tends to be more threatening. b) However, within each category this overall trend does not always hold.

We conducted post-hoc Tukey HSD tests correcting for multiple comparisons with the lsmeans() function from the lsmeans package (Lenth, 2016). Although position predicted ratings for multiple dimensions, in a few cases, such as shape, not all mean contrasts were significant (left-middle = -0.95, $p < .0001$; left-right = -1.00, p. < .0001; middle-right = -0.046, $p = .94$). For the dimension of size, not only were all the contrasts for position significant, but the difference between means was the greatest compared to all other dimensions (left-middle = -1.21, $p < .0001$; left-right = -2.73, $p < .0001$; middle-right = -1.52, $p = .001$), and this general pattern was replicated in every category (Figures 4a and 4b). We therefore conclude that, from the dimensions queried, size best indexes differences between the members of each class, and therefore is a salient and diagnostic perceptual feature.

	Left			Middle			Right		
Dimension	M	Low CI	Up CI	M	Low CI	Up CI	M	Low CI	Up CI
Approach	3.63	3.31	3.96	3.36	3.04	3.68	3.30	2.98	3.62
Arousal	3.57	3.27	3.87	4.08	3.79	4.38	4.59	4.29	4.89
Size	1.57	1.33	1.82	2.78	2.54	3.03	4.31	4.06	4.56
Shape	3.28	3.00	3.55	4.23	3.96	4.51	4.28	4.00	4.55
Speed	4.38	4.01	4.76	4.09	3.71	4.46	4.24	3.86	4.61
Threat	2.30	2.01	2.59	3.00	2.71	3.29	3.44	3.15	3.73
Valence	3.98	3.69	4.27	3.73	3.44	4.02	3.98	3.69	4.27

Table 1. Mean ratings for each dimension by position of the animal.



Figures 4a and 4b. The relationship of the fixed effect seen in a) is replicated in every category of animal in b).

Experiment 2

The previous results established that our pseudoword types are predictive of size-magnitude associations, and that size is a diagnostic feature that distinguishes between individual members of every group of animal. In this experiment, participants were tasked with naming the animals from experiment 1b using a subset of the pseudowords from experiment 1a. This task was designed to test whether the underlying iconic resemblance between these two types of stimuli would result in sound-to-meaning mappings based on the dimension of size. Thus, we predicted that the leftmost creatures, which were the smallest, would be more likely to be labelled with the small pseudowords, and the rightmost creatures, which were the biggest, would be more likely to be labelled with the big pseudowords. Thompson & Estes (2011) observed that size-sound associations were nuanced and not limited to binary, extreme contrasts. Hence, we also expected that ether type of medium pseudoword (cvCV or CVcv) would be assigned to medium-sized creatures more often than the big or small words.

Participants

Participants were sixty (mean age = 26.03, SD = 5.18, 30 men, 2 non-binary) naive monolingual English speakers with no history of speech or hearing disorders. They were compensated US$ 2.25 (average of US$ 10.19/hour).

Stimuli

We used the same the animal images from experiment 1b and 144 randomly selected pseudowords from experiment 1a: 48 small, 48 medium (24 cvCV, 24 CVcv), and 48 big (see Appendix 4 for a complete list).

Procedure

The instructions asked participants to imagine that they were alien explorers cataloguing new creatures that they had discovered, and that they were tasked with naming them. They completed 4 practice trials, reviewed the instructions, and then went on to the experimental trials. Every screen displayed 3 animals from the same category, a gray square button below each animal, and a pseudoword at the bottom of the screen. Participants had to assign that label to one of the creatures by clicking on the desired button. The smallest member of each class was always to the left, and the biggest to the right. Once again, we included attention trials. Occasionally, no pseudoword would appear onscreen, but instead there would be a sequence of 4 random digits. In that scenario, participants were supposed to pick the middle creature. Trials were pre-built such that animal class and pseudoword pairs (e.g., an image of three birds and the label keekee) were the same across participants, but the order of presentation was randomized. Each category of creature was presented 25 times: 8 times paired with small words, 8 times with medium words (4 cvCV, 4 CVcv), 8 times with big words, and once with the sequence of digits. There were 150 trials overall.

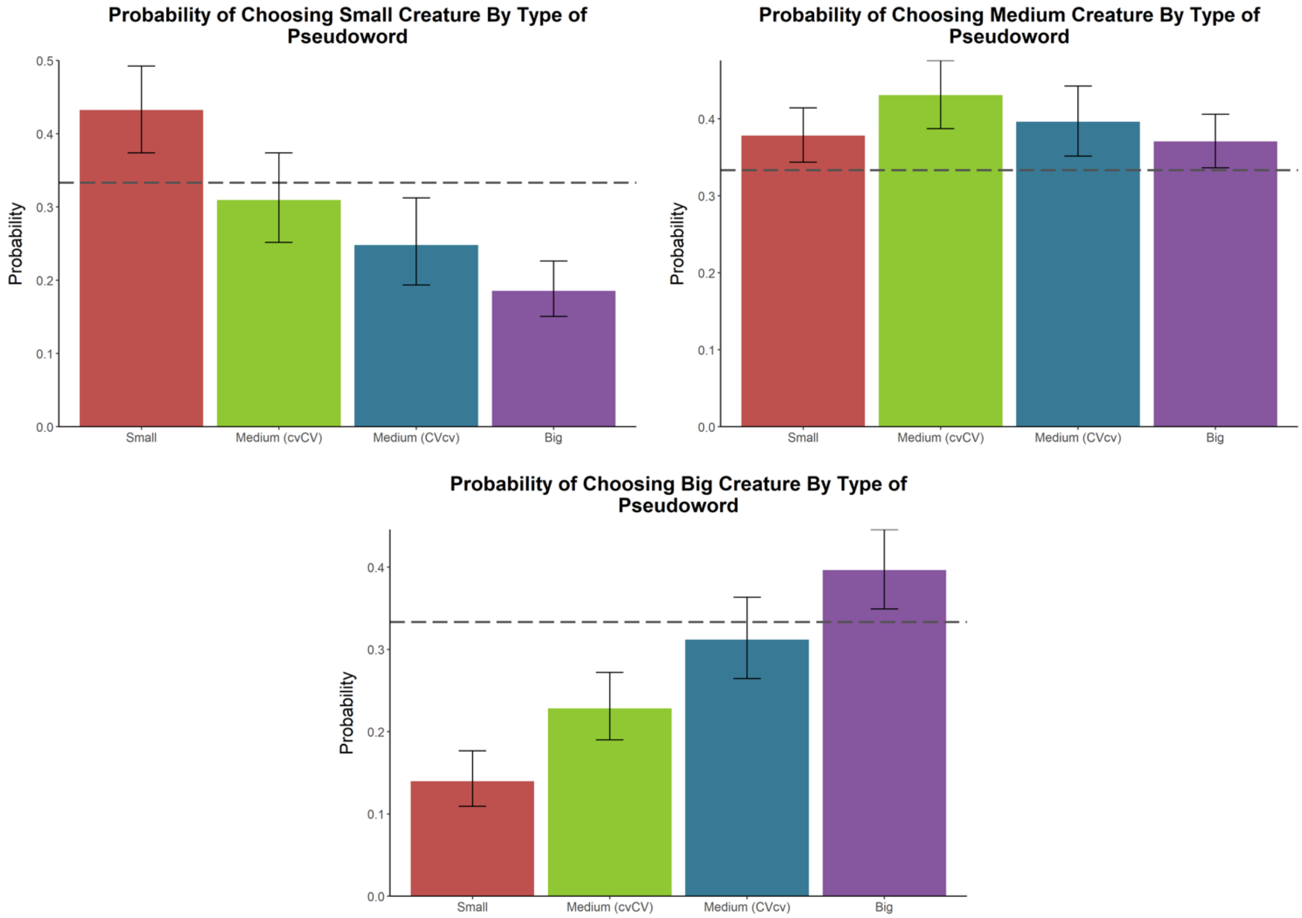
Results and Discussion

4 participants failed at least one of the attention trials, and therefore were excluded from analyses. We converted the choices of small, medium, or big animal into a binary variable by assigning it a value of 1 or 0. For example, if a small (leftmost) creature was chosen for a trial, a small_chosen factor coded this as a 1, and if it was not chosen then it was coded as a 0. Then, to predict the probability of each choice, we created 3 logistic mixed-effects models with the

glmer() function from the lme4 package in R (Bates et al., 2015). Fixed effects were the type of pseudoword and the category of animal. The interaction between these factors was not included because they induced rank deficiency within the model[2]. Random effects included intercepts and slopes for participants and intercepts for individual pseudowords. The fit of each model was analyzed with log-likelihood tests using the anova() function from the afex package (Singmann et al., 2021).

Effect of pseudoword type. Including the type of pseudoword resulted in better model fits when predicting the choice of small ($\chi^2(3) = 29.87$, $p < .0001$) and big creatures ($\chi^2(3) = 41.44$, $p < .0001$), but not the medium ones ($\chi^2(3) = 5.70$, $p = .13$). Using the emmeans() function from the emmeans package, we extracted the probability of each choice depending on the type of pseudoword that was presented (Lenth, 2021).The probability of picking a small creature was highest for small words (.43, CI[.37, .49]), followed by cvCV words (.31, CI[.25, .37]), CVcv words (.25, CI[.19, .31]), and big words (.19, CI[.15, .23]). Conversely, the probability of picking a big creature was highest for big words (.40, CI[.35, .46]), followed by CVcv words (.31, CI[.26, .36]), cvCV words (.23, CI[.19, .27]), and then small words (.14, CI[.11, .18]).

[2] In other words, the model failed to estimate interaction between some of the factor levels. This was likely due to an unbalanced number of medium pseudowords being assigned to each animal category. Though there was an equal number of cvCV and CVcv pseudowords overall, the two types of medium pseudowords were not equally represented in each category of animal.



Figures 5a, 5b, and 5c. Bars represent the probability of picking a) small, b) medium, and c) medium-sized exemplars depending on the type of pseudoword. The dashed line demarcates the chance rate.

These results indicate that participants were able to recognize affordances for iconic representation, even though the referents embodied multiple features, and consequently named the creatures based on phonological properties of the pseudowords that resembled them. Interestingly, this validates the explicit judgments from experiment 1a; the subjective size ratings can be cashed out in choice behavior in this task. Additionally, in experiment 1a, there was no significant difference in ratings between small and cvCV pseudowords, and big and CVcv pseudowords, whereas in this experiment small animals were more likely to be labelled with small pseudowords than cvCV pseudowords, and big animals were more likely to be labelled with big pseudowords than CVcv pseudowords. Considering that iconicity ratings in experiment 1a necessarily depended on conscious judgments, but in the current task participants did not need to be aware of why they were making their choices, discrepancies between the two could indicate that there are sound-to-meaning mappings which are not introspectively accessible.

It is important to highlight that there was no clear pattern of name selection for medium words; participants were not prone to choosing medium-sized creatures as referents for these words. It is possible that this pattern is a consequence of the mismatch between the number of exemplars in each category of animal, three, and the different types of pseudoword, four. That is, the two types of medium pseudowords were not mapped to the medium creatures because they indexed size magnitude with more granularity than the animals embodied. While this is a plausible explanation, the current data does not support it. For both the small and big creatures, there was a clear gradation in probability: the small words were more likely to be assigned to small exemplars, but big words were unlikely to be assigned to small creatures compared to the rest. Thus, there is a clear indication that certain pseudowords are better or worse at representing smallness. If the issue were merely that the sounds and images were not representing size at the same level of magnitude, the medium words would still be assigned to the medium creatures more often than small or big words, even if there was no difference between cvCV and CVcv words.

Instead, the interpretation more consistent with these results is that there is no sound-symbolic mapping of “medium-ness” because it is precisely a lack of distinguishing features. Of course, there is also the possibility that our pseudowords failed to capture this in-between because they were created exclusively from “big” and “small” sounds. But constructing pseudowords from “medium” sounds could be difficult methodologically because the phonemes that would enable that might not exist in a language. English has no semi-voiced consonants that could serve as a middle-ground between /b/ and /p/, for example.

Effect of animal category. A model including the effect of category was a significantly better fit for the choice of medium-sized creatures ($\chi^2(5) = 15.28$, $p = .009$), but not small ($\chi^2(5) = 4.63$, $p = .46$) or big animals ($\chi^2(5) = 7.75$, $p = .17$). Participants were not selecting small-sounding names at a higher rate for groups of animals that are, in general, tiny (e.g., butterflies), and therefore iconic mappings were made relative to the animal classes, and not in terms of absolute size. This suggests size-sound symbolism is context-specific; we cannot associate speech sounds to size without an appropriate frame of reference.

Experiment 3

Previous studies have established that we can employ prosody to communicate semantic information such as size (Nygaard et al., 2009; Tzeng et al., 2017; Tzeng et al., 2019; Perlman et al., 2021). Thus, there is good reason to expect that individuals will use iconic depiction over and above phonological features of words. For example, regardless of the specific phonemes that refer to “elephant” in any given language, speakers may lower their tone of voice, raise their volume, and extend duration to depict the creature’s large size through size-sound mapping.

Nonetheless, according to the affordance hypothesis, prosodic expression can be invited by certain speech sounds, which constitute better and worse conduits for iconicity, and may be facilitated by preexisting sound-to-meaning mappings encoded in the word form. To evaluate the hypothesis that iconicity in word form affords prosodic modulation that further augments expressiveness, we assigned pseudoword labels to various referents, in this case the creatures from the previous experiment, which either matched or mismatched in the relevant dimension of size and asked speakers to produce that each label as if it referred to the accompanying referent. For example, “teetee” was one of the pseudowords that was previously shown to be perceived as small, and reliably associated with small animals. Speakers were asked to produce this label when it was assigned to a small butterfly, in that case a match, and when it was assigned to a medium or big butterfly, a mismatch. Thus, we expected that when the sound structure of the pseudoword matched features of the referent, stronger prosodic modulation would be elicited. More precisely, we predicted that there would be an interaction between effect of the size of the animal and the type of the pseudoword, such that with increases in size magnitude, both for pseudoword and animal, speakers would produce lower f0, and higher intensity and duration.

Participants

Participants were sixty (mean age = 25.22, SD = 4.95, 30 men) naive monolingual English speakers with no history of speech or hearing disorders. They received US$ 1.50 (average of US$ 8.13/hour) for participation.

Stimuli

The images used in this study were the same as in experiment 1b. However, the category of rodents was dropped in the interest of reducing the number of trials, leaving a total of five categories. Twenty pseudowords were sampled based on the data from experiment 1. The small words were the five pseudowords with the lowest mean ratings. Medium words were five cvCV and five CVcv pseudowords that clustered closest to the general mean. Finally, the big words were the five pseudowords that had the highest mean ratings (see Appendix 5 for a complete list).

Procedure

Participants were presented with a similar story to the one from experiment 2, where they had to imagine that they were alien explorers studying new species. However, they were told that the creatures had already been named. Their task was just to catalogue the newly discovered species and record their names. Because the experiment was conducted over the internet, participants used their own recording devices to record their speech. Screens once again presented 3 members of an animal category, a red arrow above the target creature, and a pseudoword was presented at the bottom of the screen. But now there was a *Start Recording* button as well, and the pseudoword was embedded in a carrier sentence: "this creature is called a [pseudoword]." Participants had to click the recording button, speak the carrier sentence, and then produce the pseudoword. They were instructed to pronounce the words in a way that would make it easy for their fellow aliens to guess which creature they were referring to. To make within-participant comparisons between productions meaningful, the same pseudoword was paired with all 3 members of a category over the course of the task (e.g., if in one trial the leftmost spider was referred to as "gakay", the middle and the rightmost spiders would also eventually receive the same label). Thus, each one of the 20 pseudowords was repeated thrice, for a sum of 60 trials. Each of the five categories of animal was paired with 12 pseudowords: 4 of each "size" (small, small-big, big-small, big). Creature-pseudoword pairings were counterbalanced between participants through Latin-Square rotation (e.g., if for participant A, "booba" was a name for birds, for participant B, "booba" would be a name for the bugs, and so on). The individual order of trials was randomized within participants.

Results and Discussion

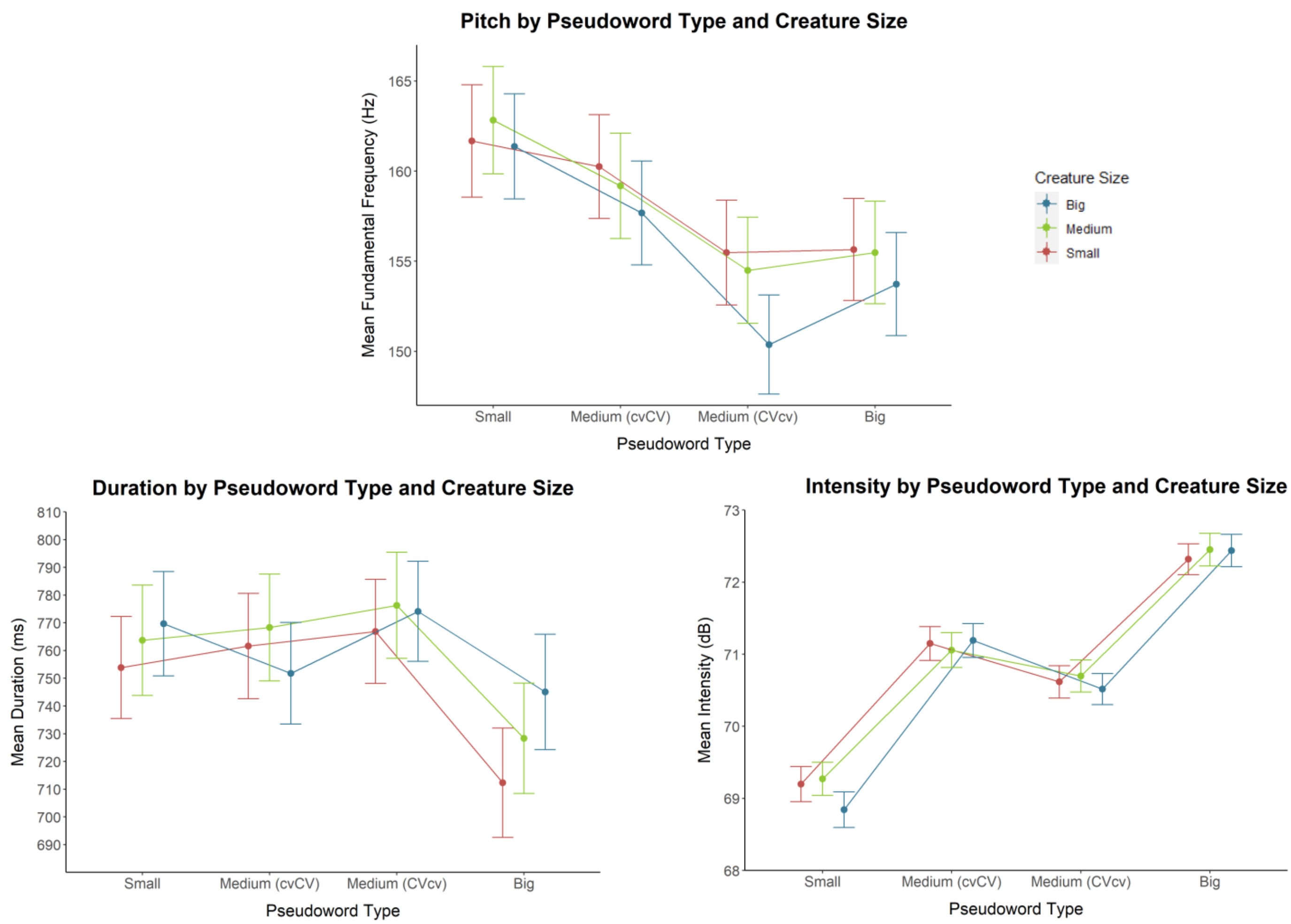
Audio preprocessing was done with Audacity® recording and editing software, version 3.13 (Audacity Team, 2021). We downsampled all audio files from 48000 Hz to 22050 Hz sampling rates, converted them from stereo to mono, and amplitude-normalized to a peak of -4.48 dB. Then, we extracted the target pseudoword from the complete sentence and edited out silence from the beginning and end of the utterances using visual cues from the waveform and audio playback. Trials were not included if there was significant issue with the production recordings (e.g., loud clicks during the production of the target pseudoword). From each segmented pseudoword utterance, we extracted the mean pitch (in Hertz), mean intensity (in decibels), and total duration (in milliseconds) using Praat, version 6.2.09 (Boersma & Weenink, 2022). A few trials were reanalyzed individually because batch analysis failed to find appropriate values.

Participants were excluded from analyses for various reasons: 1 spoke Australian English instead of American English, 4 had audio issues, and 3 failed to follow instructions. Additionally, due to the noisiness of the data—in many instances literal noise in the recordings—we excluded 80 outlier trials (z-score > 3 or < -3 in pitch, intensity, or duration). Excluding outliers did not introduce or eliminate effects. It did, however, change significance levels for an interaction. But later analyses revealed that this interaction has no important theoretical implications. Consequently, we will report only the models with outliers excluded. A total of 2992 trials were analyzed.

We built 3 mixed-effects models with duration, pitch, and intensity as predicted variables with the lmer() function from the R package lme4 (Bates et al., 2015). The models were tested for best fit with the mixed() function from the afex package (Singmann et al., 2021). The fixed effects in all cases were the type of pseudoword, the size of the exemplar, the category of animal, and interactions between these effects. Because computing random slopes for participants resulted in singular fit, the random effect structure used random intercepts only for both participants and pseudowords (Brown, 2021). Mean contrasts were computed by performing post-hoc Tukey HSD tests correcting for multiple comparisons with the lsmeans() function from the lsmeans package (Lenth, 2016).

Effects of size of creature. Fundamental frequency or pitch was significantly affected by the size of the creature ($\chi^2(2) = 6.59$, $p = .037$). Tukey HSD tests indicated that participants tended to produce speech with lower pitch when referring to the big exemplar compared to when they referred to the small exemplar (b = -2.59, CI[0.45, 4.72], p = .0461). The small-medium and medium-big contrasts were not significant (b = 0.44 CI[-1.70, 2.58] p = .91; b = 2.15 CI[0.01, 4.29], p = .12). The size of the creature did not influence either duration or intensity ($\chi^2(2) = 2.77$, $p = .25$; $\chi^2(2) = 2.41$, $p = .30$). Although the relationship was not robust, where it holds, it could indicate that participants were responding to the size of the animals and modulating their tone of voice to represent differences in magnitude.

Effects of pseudoword size. Models including the type of pseudoword as a fixed effect were a significantly better fit than models without it for intensity, duration, and pitch ($\chi^2(3) = 49.57$, $p < .0001$; $\chi^2(3) = 16.72$, $p < .0001$; $\chi^2(3) = 17.97$, $p < .0001$). Small words were produced with the lowest intensity (M = 69, CI[68.2, 69.8]), followed by medium words, cvCV and CVcv words had comparable intensity (M = 71, CI[70.2, 71.9]; M = 70.5[69.7, 71.4]), and then big words, which had the highest intensity (72.4, CI[71.5, 73.2]). Pitch was distributed in a similarly linear function, with big and CVcv words having the lowest mean frequency (M = 155, CI[144, 167]; M = 154, CI[143, 166]), cvCV words having higher frequency than these 2 (M = 160, CI[149, 171]), and small words the highest frequency (M = 164, CI[152, 175]). In terms of duration, all types of pseudoword had similar length, except for big words, which were significantly shorter than the other 3 (small-big = 41.45, p = .006, cvCV-big = 39.07, p = .01, CVcv-big = 47.56, p = .002).



Figures 6a, 6b, and 6c. Differently colored points and lines denote the different sizes of exemplar animals. Distance between them indicate the extent to which size of the referent influenced pronunciation.

Interactions. Although log-likelihood tests indicated the fixed effect of category was not significant for duration, pitch, or intensity ($\chi^2(4) = 5.92$, p = .21; $\chi^2(4) = 2.19$, p = .701; $\chi^2(4) = 3.62$, p = .46), there was a significant interaction between animal category and type of pseudoword in the case of intensity ($\chi^2(12) = 31.23$, p = .002). This was the effect that was altered due to the exclusion of outliers. With all the trials included, this interaction becomes insignificant ($\chi^2(12) = 18.56$, $p = .10$). While this could indicate that participants were pronouncing the words differently depending on the pairing between category of animal and pseudoword, a post hoc Tukey HSD test revealed no significant contrasts.

If matching animal and word size resulted in prosodic modulation, as predicted, we would observe stronger effects in some levels of the two factors. However, there was no interaction between the type of pseudoword and the size of the creatures, which means that participants were not interpreting congruency between features of the referent and features of the word form as affordances for iconic expression in prosody. But nor did participants use prosody as a disambiguating tool when word form provided no iconic mappings as was observed in Tzeng et al. (2017, 2019).

More fundamentally, the insignificant effects of duration and intensity, coupled with the very weak and inconsistent effect of f0 signal that by and large, the majority of the variation in the data reflected phonetic properties of the pseudowords, and were not a result of prosodic modulation by the participants. The other possibility is that the sound-symbolic content of the pseudowords prompted participants to alter their tone of voice to reflect it. Considering that this experiment was conducted online and involved no social interaction, the former explanation is more plausible. It is worth noting that even in this situation, it appears that participants still referred to small creatures with higher pitch than they did when referring to big creatures, which might speak to the resilience of this effect.

Because features like vowel height and voicing were constitutive of our pseudoword classes, it is likely that these phonemic properties at least contributed to the differences observed. Because the pseudowords selected for this study had the highest size ratings, they share a few striking elements. For one, they all begin with "boo", which indicates that this sound aptly represents largeness. But looking at the individual pseudowords, it appears that words that end with /a/ have quite smaller duration (Figure 7). Three out of five of the big words end in this sound. It might be that placing /a/ at the end of these words makes vowel reduction easier, and therefore results in shorter length (Fourakis, 1991). In any case, none of the effects of pseudoword type are likely to reflect purposeful prosodic modulation by the participants.

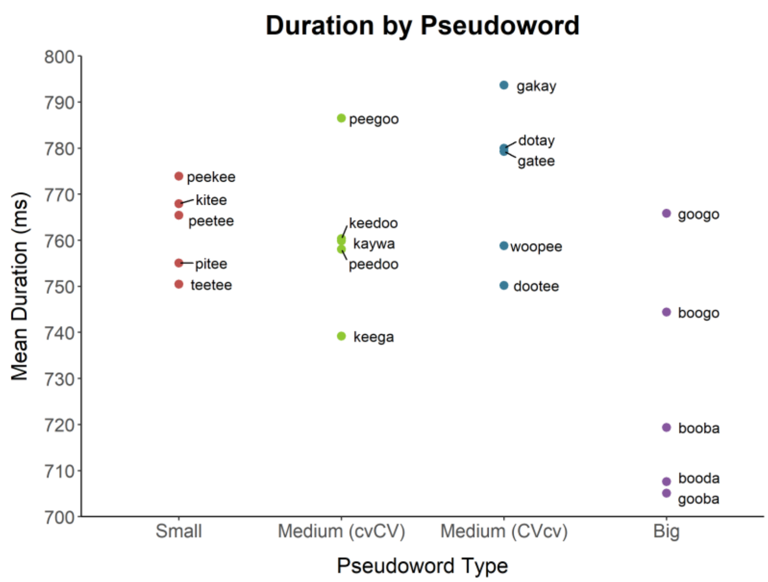


Figure 7. It appears that the words ending in /a/ were substantially shorter than the rest, with the exception of "kaywa."

General discussion

The current study explored a few different versions of the idea that sound symbolism can be understood as an affordance. Our results from experiments 1a and 1b determined that the pseudowords generated were consistently mapped to specific sizes, and that size was the feature that best helped distinguish between animals in each category. In experiment 2, participants were able to seize upon these latent iconic mappings. They disproportionally preferred names that had small-sounding phonemes for small creatures, and big-sounding phonemes for big creatures, even when they were as diverse as ostriches and Labradors. The idea that word-meaning pairings afford iconic representation when they are diagnostic of referents was largely supported by our findings. However, future studies need to test more specific and novel predictions. Perhaps a fruitful direction would be trying to forecast when sound-symbolic mappings will fail to materialize. For example, if iconicity serves to facilitate linguistic and perceptual processing, is it less likely to be used when it provides no obvious benefits for purposes? In other words, is sound symbolism used opportunistically, towards particular goals, as affordances are?

Our attempt at probing the affordance hypothesis in the context of multimodal communication is inconclusive. Participants did not exaggerate aspects of their prosody in response to specific word-referent pairings. Further, they did not use intensity or duration to convey the size of the animals, and the effect of pitch was small. It is likely that participants did not parse the circumstances of the study as a communicative context, and thus merely recited the pseudowords without prosodic modulation. Subsequent studies on this topic would benefit from placing participants in social situations, as language users are more likely to employ prosody in their speech when they are engaged in real communication (Tzeng et al., 2017, 2019).

It should be highlighted that, if the results we observed in experiment 3 reflect basic acoustic properties of the pseudowords, this is still highly informative. It implies that, at least partly, what characterized big words was low pitch and high intensity, and small words were defined by low intensity and high pitch. This is consistent, and in fact predicted by the frequency code hypothesis, which states that sounds are perceived as small or big based on their frequency (Ohala, 1984; 1994). Intensity, in turn, has been positively correlated with judgments of size magnitude in previous studies, which is also consistent with our observations in experiment 3

(Perlman et al., 2021). Finally, increase in duration has also been associated with largeness, but in this experiment the opposite was the case, as the big words had the shortest duration.

We found that our pseudowords in experiment 2 were assigned to the animals "graded function" like in Thompson and Estes' study (2011), in the sense there was a difference in choice behavior when comparing small, medium (cvCV and CVcv), and big words. Nevertheless, this did not reflect a preference on the part of participants for selecting medium pseudowords for medium-sized animals. The difference was observed simply because there was a preference in the case of small and big words, but not in the case of medium pseudowords. In their analysis, the Thompson and Estes used ANOVA to compare the average number of "big" phonemes assigned to referents of varying sizes. They demonstrated that, on average, bigger objects were paired with pseudowords that had more big-sounding phonemes. The different conclusions could stem from disparities in analytical strategies, considering that we used logistic mixed-effects models to assess our data and Thompson and Estes used ANOVA. But they could also indicate that sound symbolism can only aptly represent the extremes in a given dimension. In that case, the graded function observed by Thompson and Estes would not arise from a pattern of sound-referent responses for medium-sized objects, but a lack of pattern that pulled the average amount of big phonemes for medium objects towards the center.

The results from experiments 1a and 2 lend credibility to the idea that iconicity ratings—asking participants straightforwardly whether a word resembles its meaning—really measure iconicity (Winter & Perlman, 2021). On broad strokes, the pseudoword ratings were predictive of behavior in the animal-pseudoword matching task. But the lack of significant differences in size ratings between the small and cvCV words, and between the CVcv and big words also point to a curious discontinuity between explicit judgments and sound-to-meaning mappings. When participants chose labels for creatures varying in size, it is evident that the small words were more likely to be assigned to small creatures than the cvCV words, and the CVcv words were more likely to be assigned to big creatures than the CVcv words. Thus, it may be that the finer distinctions between these pseudowords were not consciously accessible to our participants in experiment 1a. Alternatively, perhaps a Likert-scale is not granular enough to detect them.

The serendipitous discovery that the order of syllables affects how their iconic relationships are perceived is another thread worth unraveling. CVcv words were rated as considerably bigger than cvCV words in experiment 1a. Further, the former was more likely to be chosen for big creatures and the latter for small creatures. If our interpretation that stress is driving this effect is correct, it implies that it is not word form in at a high-level of abstraction that is mapped onto perceptual features, but instead the actual speech stream with all its quirks. Regardless of the specific cause, this phenomenon demonstrates that some of the nuances of how speech sounds come to embody iconic meaning are still not well understood.

In sum, we found that size was a particularly salient dimension that characterized the individual differences between the animals in each group, and as a result participants were able to use the size magnitude evoked by the phonology of the pseudowords to name the creatures accordingly. The affordance hypothesis conceptualized as a constraint on the range of possibilities for iconic depiction, as a mechanism that facilitates linguistic processing, found support in these experiments. Though the idea that sound symbolism in word form affords iconic representation through prosody was not supported, this was likely a result of methodological issues. Therefore, this hypothesis might yet be vindicated.

Appendix

Appendix 1. Pseudowords used in experiment 1a.

Small Pseudowords

peepee	paykee	taytee	kaypee
peepay	paykay	taytay	kaypay
peetee	teekay	taykee	kaytee
peetay	tipee	taykay	kaytay
peekee	tipay	keepee	kaykee
peekay	titee	keepay	kaykay
pipee	titay	keetee	
pipay	tikee	keetay	
pitee	tikay	keekee	
pitay	taypee	keekay	
pikee	teepee	kipee	
pikay	teepay	kipay	
paypee	teetee	kitee	
paypay	teetay	kitay	
paytee	teekee	kikee	
paytay	taypay	kikay	

Big Pseudowords

baba	boodoo	dowo	gogo
babo	booga	dowoo	gogoo
baboo	boogo	dooba	gowa
bada	boogoo	doobo	gowo
bado	boowa	dooboo	gowoo
badoo	boowo	dooda	gooba
baga	boowoo	doodo	goobo
bago	daba	doodoo	gooboo
bagoo	dabo	dooga	gooda
bawa	daboo	doogo	goodo
bawo	dada	doogoo	goodoo
bawoo	dado	doowa	googa
boba	dadoo	doowo	googo
bobo	daga	doowoo	googoo
boboo	dago	gaba	goowa
boda	dagoo	gabo	goowo
bodo	dawa	gaboo	goowoo
bodoo	dawo	gada	waba
boga	dawoo	gado	wabo
bogo	doba	gadoo	waboo
bogoo	dobo	gaga	wada
bowa	doboo	gago	wado

Medium Pseudowords (CVcv)

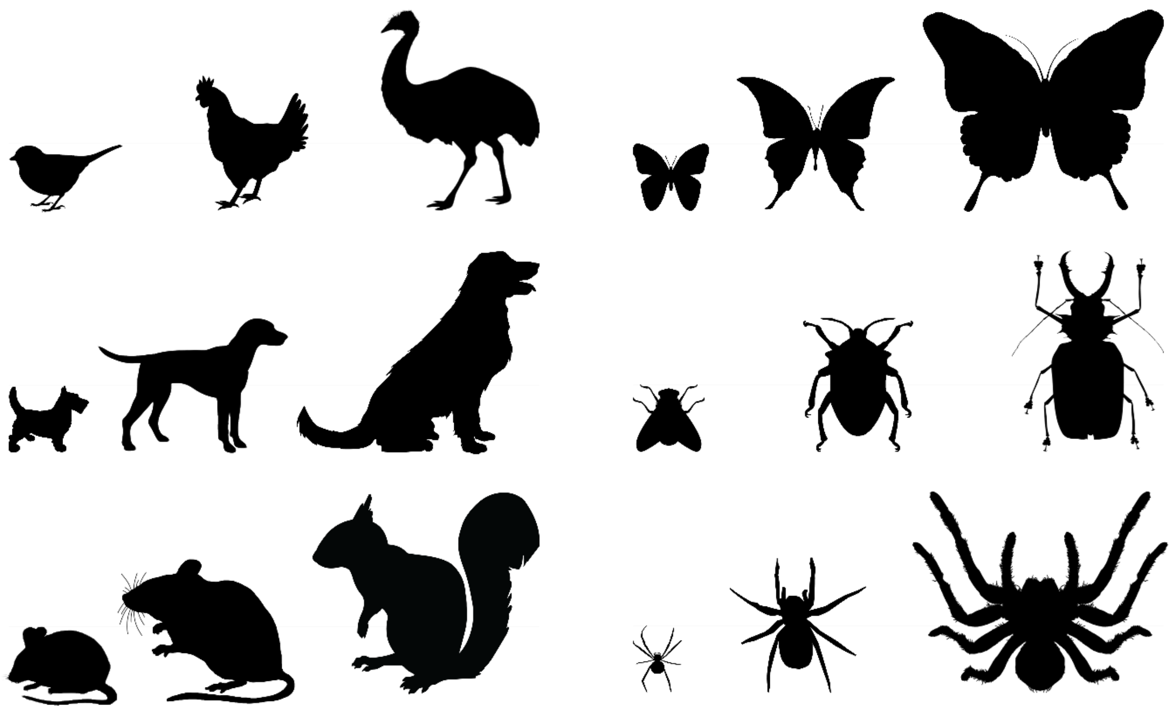
bapee	dapee	gapee	wapee
bapay	dapay	gapay	wapay
batee	datee	gatee	watee
batay	datay	gatay	watay
bakee	dakee	gakee	wakee
bakay	dakay	gakay	wakay
bopee	dopee	gopee	wopee
bopay	dopay	gopay	wopay
botee	dotee	gotee	wotee
botay	dotay	gotay	wotay
bokee	dokee	gokee	wokee
bokay	dokay	gokay	wokay
boopee	doopee	goopee	woopee
boopay	doopay	goopay	woopay
bootee	dootee	gootee	wootee
bootay	dootay	gootay	wootay
bookee	dookee	gookee	wookee
bookay	dookay	gookay	wookay

Medium Pseudowords (cvCV)

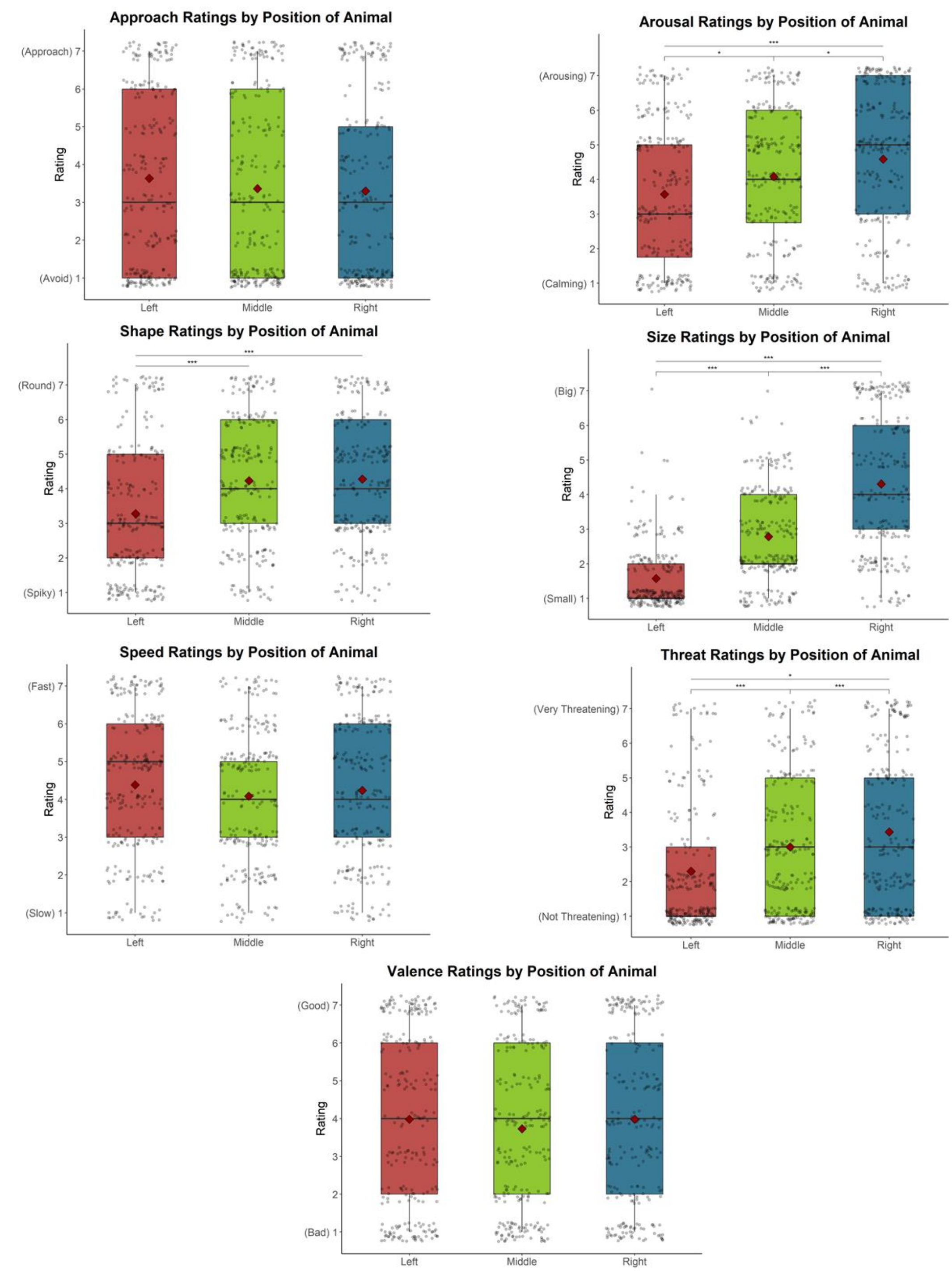
peeba	payda	tiga	keewa
peebo	paydo	tigo	keewo
peeboo	paydoo	tigoo	keewoo
peeda	payga	tiwa	kiba
peedo	paygo	tiwo	kibo
peedoo	paygoo	tiwoo	kiboo
peega	paywa	tayba	kida
peego	paywo	taybo	kido
peegoo	paywoo	tayboo	kidoo
peewa	teeba	tayda	kiga
peewo	teebo	taydo	kigo
peewoo	teeboo	taydoo	kigoo
piba	teeda	tayga	kiwa
pibo	teedo	taygo	kiwo
piboo	teedoo	taygoo	kiwoo
pida	teega	taywa	kayba
pido	teego	taywo	kaybo
pidoo	teegoo	taywoo	kayboo
piga	teewa	keeba	kayda
pigo	teewo	keebo	kaydo

bowo	doda	goba	wadoo	pigoo	teewoo	keeboo	kaydoo
bowoo	dodo	gobo	waga	piwa	tiba	keeda	kayga
booba	dodoo	goboo	wago	piwo	tibo	keedo	kaygo
boobo	doga	goda	wagoo	piwoo	tiboo	keedoo	kaygoo
booboo	dogo	godo	wawa	payba	tida	keega	kaywa
booda	dogoo	godoo	wawo	paybo	tido	keego	kaywo
boodo	dowa	goga	wawoo	payboo	tidoo	keegoo	kaywoo

Appendix 2. Images used in experiment 1b.



Appendix 3. Rating distributions from experiment 1b



Appendix 4. Pseudowords used in experiment 2.

Small Pseudowords				Big Pseudowords			
keepee	teepee	paytee	taypay	boowa	dowa	gabo	wawoo
taytay	keekee	kaytee	keetay	dabo	dawoo	booba	godoo
keetee	peepay	peepee	paytay	bago	dodoo	gooba	boobo
teekee	paykee	kipay	peekay	dawa	dooga	dago	bodoo
tikee	keekay	paypay	pipay	doodo	dogoo	bawoo	goodo
taytee	teepay	tipay	tipee	dada	wada	wawo	dadoo
peetee	pipee	taykay	teekay	dodo	gaga	baboo	wabo
kaypay	pikee	taykee	taypee	dogo	bodo	boba	dawo
paypee	titee	paykay	pitay	doogo	boodoo	godo	dowoo
kaypee	pikay	kaytay	teetee	googo	daba	baga	dowo
tikay	kaykay	titay	kikee	doda	bawo	wawa	boogo
teetay	keepay	peetay	peekee	wago	goodoo	goobo	googoo
Medium Pseudowords (cvCV)				**Medium Pseudowords (CVcv)**			
teegoo	peebo	paywoo	teewoo	bopee	gopee	gookee	wokay
keego	teedoo	kidoo	taywo	wookee	bookay	wakay	dookee
peego	kaygoo	piba	kiba	doopee	dokay	watay	gotee
paydoo	payda	teewa	kaybo	doopay	wootee	dakay	bapay
tido	peewa	teego	payga	bokay	wopay	wokee	dotay
piboo	tibo	keedoo	tayda	gootee	gootay	botee	goopee

Appendix 5. Pseudowords used in experiment 3.

Small Pseudowords	Medium Pseudowords (cvCV)	Medium Pseudowords (CVcv)	Big Pseudowords
kitee	keedoo	gakay	gooba
teetee	kaywa	gatee	booba
pitee	peedoo	dootee	boogo
peetee	peegoo	dotay	boogo
peekee	keega	woopee	googo

References

Abercrombie, D. (1982). *Elements of General Phonetics* (Paperback). Edinburgh University Press.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Audacity Team. (2021). *Audacity(R): Free Audio Editor and Recorder* (Version 3.0.0) [Computer software]. https://audacityteam.org/

Bates, D., Martin, 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

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism*. Cambridge University Press. https://doi.org/10.1017/CBO9780511751806

Berlin, B. (2006). The first congress of ethnozoological nomenclature. *Journal of the Royal Anthropological Institute*, *12*(s1), S23–S44. https://doi.org/10.1111/j.1467-9655.2006.00271.x

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

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (version 6.2.09) [Computer software]. http://www.praat.org/

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The Journal of Abnormal and Social Psychology*, *50*(3), 388–393. https://doi.org/10.1037/h0046820

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351. https://doi.org/10.1177/2515245920960351

Charalambides, N. (2021, August 24). *We recently went viral on TikTok—Here's what we learned*. Prolific. https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/S0140525X1500031X

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., … Winter, B. (2021). 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

Diffloth, G. (1994). I: Big, a: Small. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism*. Cambridge University Press. https://doi.org/10.1017/CBO9780511751806

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: A Journal of General Linguistics*, *3*(1), 4. https://doi.org/10.5334/gjgl.444

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

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of Ideophones from five languages. *Language*, *92*(2), e117–e133. https://doi.org/10.1353/lan.2016.0034

Ferreira, F., & Lowder, M. (2016). Prediction, Information Structure, and Good-Enough Language Processing. In *Psychology of Learning and Motivation* (Vol. 65). https://doi.org/10.1016/bs.plm.2016.04.002

Flaksman, M. (2017). *Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage*. https://doi.org/10.1075/ill.15.02fla

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America*, *90*(4), 1816–1827. https://doi.org/10.1121/1.401662

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Psychology Press.

Godoy, M. C., de Souza Filho, N. S., de Souza, J. G. M., França, H. A. N., & Kawahara, S. (2020). Gotta Name'em All: An Experimental Study on the Sound Symbolism of Pokémon Names in Brazilian Portuguese. *Journal of Psycholinguistic Research*, *49*(5), 717–740. https://doi.org/10.1007/s10936-019-09679-2

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X15001247

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology*, *108*(2), 229–241. https://doi.org/10.1016/j.jecp.2010.09.005

Heyman, T., Maerten, A.-S., Vankrunkelsven, H., Voorspoels, W., & Moors, P. (2019). Sound-symbolism effects in the absence of awareness: A replication study. *Psychological Science*, *30*(11), 1638–1647. https://doi.org/10.1177/0956797619875482

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 88–97.

House, A. S., & Fairbanks, G. (1953). The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113. https://doi.org/10.1121/1.1906982

Hung, S.-M., Styles, S. J., & Hsieh, P.-J. (2017). Can a word sound like a shape before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science*, *28*(3), 263–275. https://doi.org/10.1177/0956797616677313

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., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound Symbolism Facilitates Word Learning in 14-Month-Olds. *PLOS ONE*, *10*(2), e0116494. https://doi.org/10.1371/journal.pone.0116494

Klank, L. J. K., Huang, Y.-H., & Johnson, R. C. (1971). Determinants of success in matching word pairs in tests of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, *10*(2), 140–148. https://doi.org/10.1016/S0022-5371(71)80005-1

Klink, R. R. (2000). Creating Brand Names With Meaning: The Use of Sound Symbolism. *Marketing Letters*, *11*(1), 5–20. https://doi.org/10.1023/A:1008184423824

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28. https://doi.org/10.1016/j.cognition.2009.08.016

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning & Verbal Behavior*, *10*(4), 427–429. https://doi.org/10.1016/S0022-5371(71)80042-7

Lenth, R. (2016). Least-Squares Means: The {R} Package {lsmeans}. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lenth, R. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.7.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1274–1281. https://doi.org/10.1037/xlm0000235

Nielsen, A. K., & Dingemanse, M. (2020). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*(1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

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

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*. Cambridge University Press. https://doi.org/10.1017/CBO9780511751806

Ohtake, Y., & Haryu, E. (2013). Investigation of the process underpinning vowel-size correspondence. *Japanese Psychological Research*, *55*(4), 390–399.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. https://doi.org/10.1016/j.jecp.2012.05.004

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

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350. https://doi.org/10.1075/gest.14.3.03per

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*, 1433. https://doi.org/10.3389/fpsyg.2018.01433

Perlman, M., Paul, J., & Lupyan, G. (2021). Vocal communication of magnitude across language, age, and auditory experience. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001103

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*(277), 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 of London. Series B, Biological Sciences*, *369*(1651). 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), e12572. https://doi.org/10.1111/desc.12572

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

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

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

Saussure, F. de. (1983). Chapter I: Nature of the Linguistic Sign. In C. Bally, A. Sechehaye, & A. Riedlinger (Eds.), & R. Harris (Trans.), *Course in General Linguistics*. Open Court.

Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, *36*(1), 396–410. https://doi.org/10.3765/bls.v36i1.3926

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*. https://doi.org/10.3758/s13423-021-01883-3

Sidhu, D., & Pexman, P. (2017). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*(5), 1618–1643. https://doi.org/10.3758/s13423-017-1361-1

Silva, V. M., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science*, *7*(1), 182056. https://doi.org/10.1098/rsos.182056

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *afex: Analysis of Factorial Experiments* (R package version 1.0-1) [Computer software]. https://CRAN.R-project.org/package=afex

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

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

Tsur, R. (2006). Size–sound symbolism revisited. *Journal of Pragmatics*, *38*(6), 905–924. https://doi.org/10.1016/j.pragma.2005.12.002

Tsuru, S., & Fries, H. S. (1933). A problem in meaning. *The Journal of General Psychology*, *8*(1), 281–284. https://doi.org/10.1080/00221309.1933.9713186

Tzeng, C., Duan, J., Namy, L., & Nygaard, L. (2017). Prosody in speech as a source of referential information. *Language, Cognition and Neuroscience*, *33*, 1–15. https://doi.org/10.1080/23273798.2017.1391400

Tzeng, C., Nygaard, L., & Namy, L. (2017). The specificity of sound symbolic correspondences in spoken language. *Cognitive Science*, *41*(8), 2191–2220. https://doi.org/10.1111/cogs.12474

Tzeng, C. Y., Namy, L. L., & Nygaard, L. C. (2019). Communicative Context Affects Use of Referential Prosody. *Cognitive Science*, *43*(11), e12799. https://doi.org/10.1111/cogs.12799

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017). Developmental change in children's sensitivity to sound symbolism. *Journal of Experimental Child Psychology*, *160*, 107–118. https://doi.org/10.1016/j.jecp.2017.03.004

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language*, *93*(1), 10–19. https://doi.org/10.1016/j.bandl.2004.07.006

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

Wichmann, Søren, Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy*, *12*, 844–858. https://doi.org/10.3390/e12040844

Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., & Grawunder, S. (2021). Rethinking the frequency code: A meta-analytic review of the role of acoustic body size in communicative phenomena. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840), 20200400. https://doi.org/10.1098/rstb.2020.0400

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